



सत्यमेव जयते

INDIAN AGRICULTURAL
RESEARCH INSTITUTE, NEW DELHI

22884
417

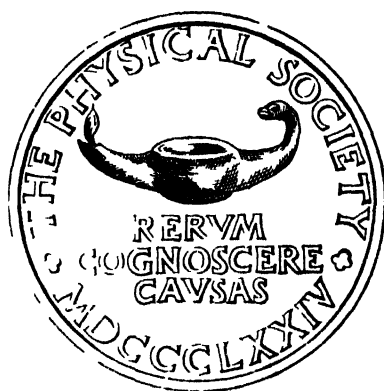
I.A.R.I.6.

GIP NLK—B-3 I.A.R.I.—10-5 55—18,000

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

FROM JANUARY 1948 TO JUNE 1948

VOLUME 60



Published by
THE PHYSICAL SOCIETY
1 Lowther Gardens, Prince Consort Road,
London S.W. 7

Printed by
TAYLOR AND FRANCIS, LTD.,
Red Lion Court, Fleet Street, London E.C. 4

CONTENTS

Part 1. 1 January 1948

	PAGE
D. C. PACK, (Dr.) W. M. EVANS and H. J. JAMES. The propagation of shock waves in steel and lead	1
ROBERT WEIL. The variation with temperature of metallic reflectivity	8
D. A. WRIGHT. Work function and energy levels in insulators	13
D. A. WRIGHT. Energy levels in oxide cathode coatings	22
(Dr.) N. R. CAMPBELL and (Dr.) L. HARTSHORN. The experimental basis of electro-magnetism : Part II - Electrostatics	27
A. W. BREWER, (Dr.) B. C'WILONG and (Dr.) G. M. B. DOBSON. Measurement of absolute humidity in extremely dry air	52
C. H. COLLIF, J. B. HASTED and D. M. RITSON. The cavity resonator method of measuring the dielectric constants of polar liquids in the centimetre band . .	71
(Dr.) B. BLEANEY and R. P. PENROSE. Pressure broadening of the inversion spectrum of ammonia : Part II—Disturbance of thermal equilibrium at low pressures . .	83
(Miss) DOROTHY G. FISHER. The molecular structure and arrangement in stretched natural rubber	99
Reviews of books	114

Part 2. 1 February 1948

(Dr.) H. WILMAN. The structure of photo-sensitive lead sulphide and lead selenide deposits and the effect of sensitization by oxygen	117
(Dr.) W. A. PROWSE. Solid diagrams illustrating resonance phenomena	132
(Dr.) L. R. G. TRELOAR. Stresses and birefringence in rubber subjected to general homogeneous strain	135
C. H. COLLIE, J. B. HASTED and D. M. RITSON. The dielectric properties of water and heavy water	145
(Dr.) KUN HUANG. Quantum mechanical calculation of the heat of solution and residual resistance of gold in silver	161
(Dr. W. E. DUNCANSON and (Prof.) C. A. COULSON. Electron momenta in atoms . .	175
(Dr.) MOHAMMED CHAUDHRI and A. G. FENTON. Some experiments with adjustable Geiger-Müller counters	183
N. F. ASTBURY. The calibration of hydrophones and crystal transceivers	193
N. CORCORAN and J. M. HOUGH. A method of computing a vertical section of the combined polar diagram of a radio aerial, a flat earth and a vertical screen . .	203
Letters to the Editor :	
(Dr.) F. C. FRANK. An isotopic abundance rule and its bearing on the origin of the nuclei	211
Reviews of books	212
Instructions to Authors	insertion

Part 3. 1 March 1948

(Prof.) L. F. BATES and E. G. HARRISON. The adiabatic temperature changes accompanying the magnetization of some ferromagnetic alloys in low and moderate fields	213
--	-----

	PAGE
(Prof.) L. F. BATES and E. G. HARRISON. The adiabatic temperature changes accompanying magnetization in low and moderate fields : A further study of iron	225
R. STREET. The variation with magnetization of Young's modulus for cobalt	236
K. W. PLESSNER. The electric strength of dielectric films	243
(Prof.) C. A. COULSON. Excited electronic levels in conjugated molecules : I. Long wavelength ultra-violet absorption of naphthalene, anthracene and homologs	257
(Dr.) H. C. LONGUET-HIGGINS. Excited electronic levels in conjugated molecules : II. On the symmetry and multiplicity of molecular states	270
J. B. BIRKS. The measurement of the permeability of low-conductivity ferromagnetic materials at centimetre wavelengths	282
T. SMITH. On perfect optical instruments	293
(Prof.) E. N. da C. ANDRADE. A new device for maintaining constant stress in a rod undergoing plastic extension	304
Letters to the Editor :	
(Prof.) L. F. BATES and J. H. DAVIS. The effect of temperature on the heat changes accompanying magnetization of the nickel-silicon alloy W.5	307
Corrigenda	308
Reviews of books	308

Part 4. 1 April 1948

W. E. MOFFITT and (Prof.) C. A. COULSON. The electronic structure and bond lengths of coronene and pyrene	309
(Miss) C. C. DILWORTH. The influence of surface films on the electrical behaviour of contacts	315
B. DONOVAN. The specific heat of a linear ionic lattice	325
(Dr.) H. WILMAN. The interpretation and application of electron-diffraction 'Kikuchi-line' patterns—Part I. The determination of the crystal unit cell, its orientation and the crystal symmetry	341
(Dr.) E. P. WOHLFARTH. Thermionic emission constants and band overlap	360
(Dr.) H. D. BUSH and (Dr.) R. S. TEBBLE. The Barkhausen effect	370
R. E. BELIN. A radiosonde method for atmospheric potential gradient measurements	381
(Miss) MARJORIE WILLIAMSON. A note on the Lecher wire method of measuring impedance	388
(Prof.) N. F. MOTT. Slip at grain boundaries and grain growth in metals	391
Letters to the Editor :	
(Dr.) B. BLEANEY and (Dr.) R. P. PENROSE. Paramagnetic resonance at low temperatures in chromic alum	395
J. R. HASTED. Milli-microsecond pulse generation by electron bunching.	397
Reviews of books	398

Part 5. 1 May 1948

(Dr.) G. L. PICKARD and (Prof.) F. E. SIMON. A quantitative study of the expansion method for liquefying helium	405
(Dr.) M. DÉSIRANT and (Dr.) D. SHOENBERG. Penetration of magnetic field into superconductors : I. Measurements on thin cylinders	413
(Dr.) K. D. FROOMI. The rate of growth of current and the behaviour of the cathode spot in transient arc discharges	424

	PAGE
(Dr.) S. P. SINHA. Blue and ultra-violet bands of K_2	436
(Dr.) S. P. SINHA. Ultra-violet bands of Li_2	443
(Dr.) S. P. SINHA. Ultra-violet bands of NaK	447
R. M. SHILLITO. An extension of Kapitza's theory of δ -radiation	453
G. K. HORTON. Angular distribution in internal pair creation	457
(Dr.) E. BRODA, (Dr.) W. E. GRUMMIF, J. GUÉRON, L. KOWARSKI and (Dr.) G. WILKINSON. The correction for self-weakening in β -ray measurements	460
D. G. E. MARTIN, (Dr.) H. O. W. RICHARDSON and YUN-KUFI HSÜ. The β -ray spectrum of $ThC''D$	466
(Prof.) B. FERRIET and (Dr.) M. KROOK. On the solution of scattering and related problems	481
(Dr.) A. C. B. LOVELL and J. A. CLEGG. Characteristics of radio echoes from meteor trails : I. The intensity of the radio reflections and electron density in the trails	491
Letters to the Editor :	
(Dr.) A. C. MENZIES and J. SKINNER. Polarization of second order Raman effect in alkali halides	498
Reviews of books	499

Part 6. 1 June 1948

(Miss) B. ZAJAC, (Dr.) E. BRODA and (Prof.) N. FATHER. A further study of the γ -radiation from polonium	501
(Mrs.) PHYLLIS NICOLSON and (Dr.) V. SARABHAI. The semi-diurnal variation in cosmic ray intensity	509
W. M. GIBSON and D. L. LIVESLY. The neutrons emitted in the disintegration of nitrogen by deuterons	523
L. GFRÖ and R. SCHMID. Dissociation energy of the NO molecule	533
(Dr.) B. BLFANEY and (Dr.) R. P. PENROSE. Collision broadening of the inversion spectrum of ammonia : III. The collision cross-sections for self-broadening and for mixtures with non-polar gases	540
(Dr.) M. BLACKMAN and (Dr.) J. L. MICHIELS. Efficiency of counting systems	549
(Prof.) L. B. LOEB. Certain aspects of the mechanism of spark discharge	561
(Dr.) G. F. J. GARLICK and A. F. GIBSON. The electron trap mechanism of luminescence in sulphide and silicate phosphors	574
N. F. ASHBURY. The moving coil galvanometer as a circuit element	590
Letters to the Editor :	
W. MOFFITT. The structure of carbon monoxide	597
N. DAVY. A note on the shape of the polepieces of a synchrotron magnet	598
Reviews of books	599
Index to Volume 60	601
Index to Reviews of books, Volume 60	606

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 60, PART 1

1 January 1948

No. 337

The Propagation of Shock Waves in Steel and Lead

By D. C. PACK, W. M. EVANS AND H. J. JAMES

Armament Research Department, Ministry of Supply

Communicated by N. F. Mott ; MS. received 18 April 1947

ABSTRACT. An investigation is made of the stress system set up by an explosive detonating in contact with a metal surface. An extrapolation from data on the compressibility of steel and lead leads to the conclusion that the shock wave set up by the detonation has an initial velocity in steel which is less, and an initial velocity in lead which is greater, than the velocity of plane elastic waves. The time taken by the fastest pulse to penetrate various lengths of steel and of lead has been measured experimentally, and the results confirm that the plane elastic waves move more quickly for steel ; while for lead the shock wave before damping has a velocity well in excess of that of the elastic waves.

§1. INTRODUCTION

THE propagation of shock waves in metals is a subject which appears to have been but little studied from a fundamental point of view. The behaviour of a shock wave in a metallic medium will depend upon the compressibility of the medium, which in turn depends upon the type of deformation which the medium experiences. The problem is, then, in general, a very difficult one, but it is possible to obtain useful information by restriction to the simpler types of deformation, which permit a simple interpretation. It is clear that the problem is one with a number of important applications, not least in connection with detonation, by which very severe shock waves are set in motion ; the movement of the metal tube confining a charge is an obvious example, since it can control the rate at which detonation proceeds in the explosive column.

A series of experiments described below has been carried out with the object of measuring the velocity of the fastest pulse induced in steel and in lead respectively, by detonating a charge in contact with the metal. By plotting the time required for passage through various thicknesses, the initial velocity of the pulse may be found. A study of the stress system set up by the detonation, together with an extrapolation from data published by Bridgman (1931, 1940) on the compressibility of steel and lead, enables a theoretical evaluation to be made of the initial shock wave velocity in each medium. It is found in the case of steel that this velocity is rather less than that of the fastest elastic waves ; the latter are, in fact, recorded in the experiments. On the other hand, for lead, the shock wave velocity is greater than the velocity of the elastic waves, and corresponds to the velocity measured. Due to the damping of this shock wave it finally reduces to an elastic wave, and the corresponding velocity of propagation is indeed measured for the greatest thicknesses of lead used

§ 2. THE VELOCITIES OF WAVES IN METALS

Before proceeding, it may be of use to recall the nature of the different wave velocities which may be measured in metals.

For elastic disturbances in which the displacement is wholly in the direction of propagation of the waves, the waves are said to be "plane", and the velocity is given by

$$V_1 = \sqrt{\left\{ \frac{3K}{\rho} \cdot \frac{(1-\nu)}{(1+\nu)} \right\}},$$

where K is the bulk modulus of compression, ρ the density, and ν is Poisson's ratio. This is also the general "velocity of sound" for waves of dilation, involving neither shear nor rotation, and is not to be confused with the "velocity of sound" in experiments on rods, in which lateral displacement is freely allowed, and in which, on the assumption that a purely longitudinal stress may be propagated along a rod without accompanying transverse stresses, the velocity of sound is $V_2 = \sqrt{(E/\rho)}$, where E is Young's modulus. The velocity V_2 and the velocity of transverse waves $V_3 = \sqrt{(G/\rho)}$, where G is the modulus of rigidity, are not relevant to the problems under discussion in the present work, in which the conditions are such that we avoid the necessity of considering the transverse displacements (see § 3).

When the stress is at the yield point of the medium, "plastic waves" similar in nature to the plane waves, are propagated with velocity $V_4 = \sqrt{(K/\rho)}$. Since Poisson's ratio cannot exceed $\frac{1}{2}$, a result which is deducible from the relations between the elastic constants (see Southwell 1941), it follows at once that $V_4 \leq V_1$, and we may say that for all metals the plastic waves are initially slower than the plane elastic waves. However, for high amplitudes of pressure in the applied pulse, the plastic waves are capable of achieving very high velocities, if, as strain increases, the rate of change of stress with strain also increases; in other words, if the effective modulus of compression increases with strain. The higher stresses are propagated with ever increasing velocities, and so lead to a shock wave, whose velocity eventually exceeds the velocity of the elastic waves.

The waves considered above (other than shock waves) have been calculated to have velocities given (approximately) in table 1.

Table 1

	$\frac{1}{K}$ (dynes/cm ²)	ρ (gms./c.c.)		V_1 (metres/ sec.)	V_2 (metres/sec.)	V_3 (metres/ sec.)	V_4 (metres/ sec.)
Iron	0.61×10^{-12}	7.8	0.28	5950	5000-5500	3200	4600
Lead	2.20×10^{-12}	11.4	0.45	2150	1200	700	2000

The difference between the velocity of plane elastic waves and of elastic waves in rods is clearly seen in the above table, particularly for lead, and is due to the large value of Poisson's ratio for this metal. For average values of ν (say $\nu = 0.3$) the difference between V_1 and V_2 is not sufficiently marked to make the velocities experimentally distinguishable.

§3. THE STRESS SYSTEM

When an impulsive force such as an explosion occurs over a surface, the condition immediately beneath the surface is virtually one in which lateral strain is absent. This is because it is possible to choose a small thickness of the block such that the time in which the pulse from the explosion traverses this thickness is not sufficient to allow the lateral displacement which is necessary to produce lateral strain to take place across the much greater area of the impulse. The limiting depth for which the condition of no lateral strain may be assumed depends, therefore, upon the area over which the explosion occurs; and we see that we may make this assumption for a block of which the thickness is small compared with the area of the explosion. Also, if, as in our experiments, the explosion occurs over an area which is less than the whole area of block surface, the consideration above assures that there will be no interference near the axis from waves which have been reflected from the lateral boundary of the block. Under these conditions the principal axes of strain are along the axis of the block and in the plane normal to it. Denoting the principal strains by $\epsilon_1, \epsilon_2, \epsilon_3$, and the corresponding principal stresses by $\sigma_1, \sigma_2, \sigma_3$, the suffix unity referring to the axial component, we have $\epsilon_2 = \epsilon_3 = 0$, $\sigma_2 = \sigma_3$, and $\sigma_1 = p$, the uniform pressure. The components σ_2, σ_3 are set up as a direct result of the effective prevention of lateral strain by inertia, in contrast with the case of a thin bar under compression, where these components vanish on account of the freedom for expansion.

R. von Mises has given a condition for the onset of plastic flow (see, for example, Nadai 1931), viz. :—

$$(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2 = 2Y^2,$$

where Y is the yield stress in simple tension.

The left hand side of this expression is proportional to the elastic energy per unit volume used in changing the shape, i.e. the potential energy of distortion, as opposed to that causing volume change under elastic strains. The condition therefore states that plastic flow begins when the potential energy of distortion reaches a limiting value, and that this energy remains constant *during* the plastic flow.

Putting $\sigma_2 = \sigma_3$, we see that for the conditions under discussion, the equation reduces to

$$|\sigma_1 - \sigma_2| = Y.$$

The volume change in the metal is due to that part of the stress system which corresponds to a hydrostatic pressure σ equal to the mean of the applied stresses, i.e. $\sigma = \frac{1}{3}(\sigma_1 + \sigma_2 + \sigma_3) = \sigma_1 - \frac{2}{3}Y$; it is clear that if the pressure applied is so large that Y may be neglected in comparison, the change of density may be put equal to that due to a hydrostatic compression of magnitude equal to the applied pressure. Shear distortion and plastic flow do not contribute to this effect.

Let the volume of the metal under applied pressure p , be changed from V_0 to V , the decrement being denoted by $-\Delta V$.

Let us also write

$$p = \alpha \left\{ \exp \left(\beta \frac{\sqrt[3]{V_0} - \sqrt[3]{V}}{\sqrt[3]{V_0}} \right) - 1 \right\} / \left(\frac{V}{V_0} \right)^{2/3}. \quad \dots\dots(1)$$

This tentative law is suggested by quantum mechanical considerations, since it requires the pressure to increase exponentially with the change in the inter-atomic distance, and has been constructed so that $V = V_0$ when $p = 0$, and $V \rightarrow 0$ as $p \rightarrow \infty$. α, β are two parameters, by means of which an extrapolation to high pressures can be made from the results determined in experiments on volume changes due to hydrostatic compression at lower pressures. The extrapolation involved is from 30 000 atmospheres for steel, 12 000 atmospheres for lead to (approx.) 280 000 atmospheres at the surface of a target.

To determine the values of α, β we expand (1), substituting a power series in p for $-\Delta V/V_0$. To the second degree in p we obtain:—

$$-\frac{\Delta V}{V_0} = \frac{3}{\alpha\beta}p - \frac{3}{2\alpha^2\beta^2}(\beta + 6)p^2. \quad \dots\dots(2)$$

This may be compared with the empirical formulae deduced by Bridgman from experimental results on many different metals, and written in the form

$$-\frac{\Delta V}{V_0} = Ap - Bp^2.$$

In what follows, the variability of A, B with temperature is not taken into account.

For iron (Bridgman 1940), we use $A = 5.94 \times 10^{-13}$, $B = 0.60 \times 10^{-24}$ expressed in c.g.s. units. These figures lead to the values

$$\alpha = 120.16 \times 10^{10} \text{ dynes/sq. cm.}, \quad \beta = 4.203.$$

For lead (Bridgman 1931), we use $A = 24.50 \times 10^{-13}$, $B = 18.16 \times 10^{-24}$ expressed in c.g.s. units. from which we obtain

$$\alpha = 10.08 \times 10^{10} \text{ dynes/sq. cm.}, \quad \beta = 12.152.$$

It may be remarked in passing that the values A, B used above lead to values of the thermal expansion coefficient which are in good agreement with those predicted theoretically by Mott and Jones (1936); the original value of B for iron, quoted by Bridgman (1931), appeared to be too large, and has, in fact, been corrected to that used in the above calculations, as the result of experiments at higher pressures than had previously been attainable (see Bridgman 1940).

§ 4 THE SHOCK-WAVE VELOCITY

We now calculate the velocity at which a shock wave advances into a metal, using the relationship between pressure and density which we have already postulated in equation (1).

Suppose the shock wave moves with a velocity U into a metal at rest. Let the velocity of the metal after passage of the wave be u in the same direction, and let its density be changed from ρ_0 to ρ . Let the pressure jump be p .

Then the equation of continuity is

$$\rho(U - u) = \rho_0 U, \quad \dots\dots(3)$$

$$\text{and the momentum equation is} \quad p = \rho_0 U u, \quad \dots\dots(4)$$

$$\text{from which we obtain} \quad U^2 = \frac{p}{\rho_0} \cdot \frac{p}{\rho - \rho_0}. \quad \dots\dots(5)$$

Equation (1) supplies the necessary additional information to correlate the density change with the pressure.

Corresponding to a given pressure p_0 in the detonation wave front, there will be a pressure p at the surface of the target, determined by the necessity for the pressure and mass-velocity of the metal to be equal to the pressure and mass-velocity in the explosion gases, after the propagation, from the interface, of a shock wave forwards into the metal, and backwards into the gases. The gas relationships for various high explosives have been worked out by Dr. H. Jones, and are shortly to be published. Dr. A. F. Devonshire kindly supplied us with tables based upon the relationships for the explosive used in our experiments; by assuming a pressure in the detonation wave front of 17×10^{10} dynes/sq. cm., i.e. 1100 tons/sq. in., we were able to deduce that:—

at the surface of a steel target, $p = 28.3 \times 10^{10}$ dynes/sq. cm.

at the surface of a lead target, $p = 27.1 \times 10^{10}$ dynes/sq. cm.

From these we obtain :

for steel $U = 5240$ m./sec. and $u = 690$ m./sec. ;

for lead $U = 3020$ m./sec. and $u = 790$ m./sec.

These examples are interesting since they indicate two distinct possibilities. The velocity of the shock wave in steel is seen to be less than the velocity of the plane elastic waves, which therefore travel in advance of the shock wave; in lead, however, the shock wave has an initial velocity well in excess of the velocity of the elastic waves, and is therefore the fastest pulse in this metal.

§ 5. EXPERIMENTAL METHODS

To examine experimentally the velocities of the fastest pulses, charges of explosive were detonated in contact with bars of steel and of lead. We see from what has been written above (§ 2 and § 4), that steel and lead may be expected to provide excellent examples of two different phenomena. For steel the initial shock wave velocity has been calculated to be less than that of the fastest elastic waves, while for lead it was found to move at almost $1\frac{1}{2}$ times the velocity of these plane waves, under the pressures set up by the explosive.

The charges used were made up of 45 gm. explosive, confined in copper tubing 1, 10 in. wall thickness and $1\frac{1}{4}$ in. internal diameter. The front of each charge was flat and unconfined, the rear being in the form of an 80° cone tamped with 15 g. plasticene and initiated by a No. 8 Briska detonator. The targets were 3 in. in diameter, and were of various lengths. The steel was cut from a bar of about 0.3% carbon and about 35 tons/sq. in. ultimate stress. The lead bars were carefully cast and then cut to the required lengths.

A diagram of the arrangement employed for the velocity measurements is given in figure 1. The front of the charge is in contact and central with one face of the target. Central with the other face is a $3/16$ in. polished steel ball soldered to the end of a screw which passes through a wooden support. The screw was turned until electrical contact was just not made between the ball and the target. This meant an air gap of a fraction of a thousandth of an inch between ball and face.

Through the copper tubing and near to the front of the charge is a hole about 0.1 in. diameter. In very close proximity to and central with this hole is an ionization head which consists merely of two enamelled copper wires 26 s.w.g. running side by side, separated from each other by their insulation, but with their

ends bare. Thus, when the ionized gases reach the hole in the tube, the "head" passes a current, and in this way, subject to an error to be discussed later, we are able to record the arrival of the detonation wave at the face of the target. A pulse of compression is propagated through the target, setting the metal in motion as it progresses. The approximate instant it reaches the rear of the target is recorded by the movement of the rear face, making contact with the steel ball.

The apparatus used to register the time difference between head and ball events was a Baird microsecond chronometer.

§ 6 ERRORS INVOLVED IN THE EXPERIMENTS: CORRECTIONS APPLIED

(i) The input circuits of the contour were first adjusted to trip at the same amplitude (6 v.) of voltage pulse. To overcome tripping instability caused by the long leads to the bomb chamber, a voltage of 66 v. was applied, attenuated by a 10:1 (5000, 500- Ω resistors) potentiometer. Low capacity and widely spaced event leads diminished distortion and lag of the pulses (see figure 1).

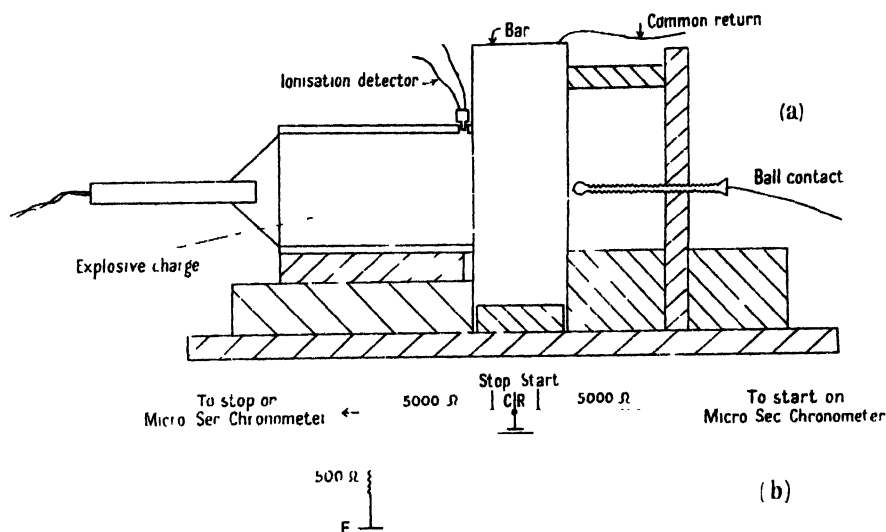


Figure 1.

(ii) The difference of pulse delays in the two leads was measured by transmitting the same pulse down both leads and measuring the delay between their arrivals at the other ends. This time interval was of the order of $0.1 \mu \text{ sec.}$, the "stop" event lead being the quicker, implying that time readings with these leads were $0.1 \mu \text{ sec.}$ too low.

(iii) The delay between the impact of detonation wave on the steel surface and the emergence of ionized gases through the hole on the periphery of the charge (normal start event) was determined by placing another ionization detector at the end of the charge (see figure 1). The delay between the two detectors was $0.4 \mu \text{ sec.}$ which added to (ii) gave a total error of $0.5 \mu \text{ sec.}$

(iv) The time-lag involved in the ball event at the end of the metal specimen (figure 1) is determined mainly by the distance between this surface and the ball.

This lag was determined experimentally by finding the delay between two balls at different known distances from the back surface. It was concluded that the maximum error derivable from this source is not greater than 10^{-7} sec.

These errors were taken into account and it is considered that the values of the times obtained are correct to $\pm 1\%$.

After making the corrections outlined above, the times of propagation of the fastest pulse in various lengths of lead and of steel have been plotted against the lengths. The gradient of a curve at a given length measures the velocity of the fastest pulse at an equal distance from the surface of the target (figures 2 and 3).

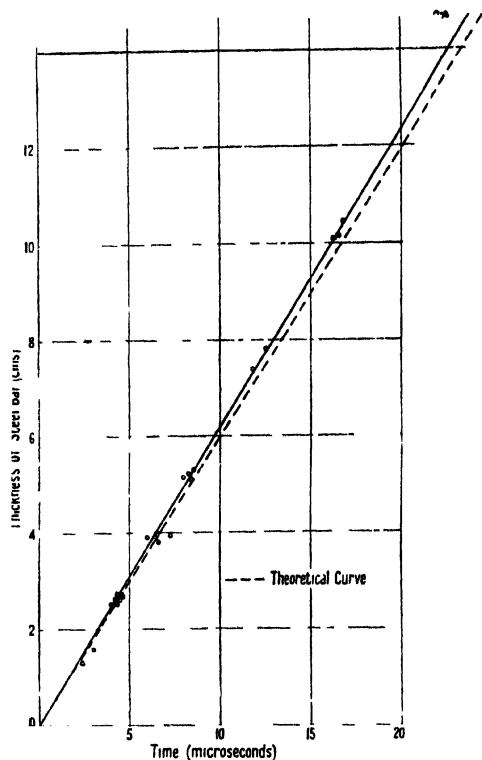


FIGURE 2.

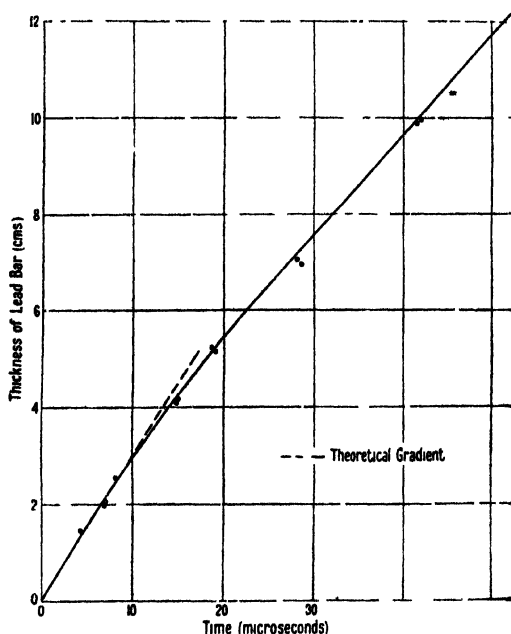


Figure 3.

§ 7. DISCUSSION OF RESULTS

The experimental values obtained for the times for the fastest wave to traverse the bars of differing lengths are plotted in figures 2 and 3. Figure 2 gives the graph for steel, and it is seen from the experimental curve that there is no appreciable decay in velocity for lengths up to at least 12 in. Since the theoretical value of the shock wave velocity is below that given by V_1 (the general velocity of elastic waves), the theoretical curve has been drawn with gradient corresponding to V_1 . The agreement with the experimental curve is very good, so that there can be little doubt that the velocity measured is that of the elastic waves, V_1 .

A different picture is presented for the bars of lead. In figure 3 the curve through the experimental points has a pronounced curvature near the origin, becoming straight as the thickness of the bar is increased. The shock wave

velocity, determined theoretically is approximately 3000 m./sec., well in excess of the velocity of elastic waves. A dashed line through the origin represents the gradient associated with the former velocity, and it is seen to lie along the experimental curve at the origin. Since, by extrapolation of the curve to zero length of bar, we obtain the undamped velocity of the shock wave, we should, in fact, expect the value at the origin to correspond to that calculated. For the longest bars of lead used, the gradient of the experimental curve leads to a velocity in close agreement with that of plane elastic waves, viz. 2100 m./sec.

We conclude then, that the pressure set up by the detonation in steel is not sufficient to propagate a shock wave at a velocity greater than that of elastic waves; in lead the shock wave is propagated much faster than elastic waves but damping occurs very rapidly with increasing length of bar, the velocity finally being that of elastic waves. It is to be emphasized that this ultimate velocity in lead is V_1 , the general velocity of elastic waves, and not $V_2 = \sqrt{(E/\rho)}$, the velocity of elastic waves associated with thin rods.

ACKNOWLEDGMENTS

The authors wish to acknowledge the kind advice and encouragement given to them by Professor N. F. Mott, F.R.S., during these investigations. They are also indebted to the Chief Scientific Officer, Ministry of Supply, for permission to publish this work.

REFERENCES

- BRIDGMAN, P. W., 1931 *The Physics of High Pressures* (London).
 BRIDGMAN, P. W., 1940, *Phys. Rev.*, **57**, 235, 273 and 342.
 MOTT, N. F., and JONES, H., 1936, *Properties of Metals and Alloys* (Oxford).
 NADAI, A., 1931, *Plasticity* (New York and London).
 SOUTHWELL, R., 1941, *Theory of Elasticity*, 2nd edition, p. 124 (Oxford).

The Variation with Temperature of Metallic Reflectivity

By ROBERT WEIL

South-West Essex Technical College

MS. received 22 May 1947

ABSTRACT. It is shown that, assuming free electrons alone to be responsible for the mechanism of metallic reflection, there is a wavelength at which the temperature variation of reflectivity is zero. Since this wavelength is directly proportional to the time of relaxation of the electrons it is to be concluded that it varies with the temperature. The effect of the bound electrons is also considered.

§ 1. INTRODUCTION

IN a recent communication (1947) it was pointed out by Price that the temperature variation of the emissivities of several metals is such as to suggest critical wavelengths, specific to each metal, for which the temperature coefficient of emissivity is zero. These wavelengths were called X-points. Several explanations were advanced but it would appear to the present writer

that these X-points can be predicted if dealt with from the point of view of reflectivity, and that, as a result of such an investigation, Price's conclusions would have to be scrutinized more closely.

§ 2. THEORY

(a) *Free electrons*

While the temperature coefficient of emissivity is defined by

$$\beta = \frac{1}{E} \cdot \frac{dE}{dT}, \quad \dots\dots(1)$$

it is desirable to consider the gradient dE/dT only, because this is simply related to the corresponding reflectivity gradient by the expression

$$\frac{dE}{dT} = - \frac{dR}{dT}. \quad \dots\dots(2)$$

Obviously β will be zero when dE/dT is zero, and, therefore, it will be in order to consider the wavelength at which dR/dT is zero: this, then, will be an X-point.

Electro-magnetic theory leads to the following expressions connecting the reflectivity R , the refractive index n , the extinction coefficient k , the conductivity σ , and the dielectric constant ϵ with the frequency of the radiation ν :

$$nk = \frac{\sigma}{\nu}, \quad \dots\dots(3)$$

$$n^2 - k^2 = \epsilon. \quad \dots\dots(4)$$

$$R = 1 - \frac{4n}{(n+1)^2 + k^2}. \quad \dots\dots(5)$$

It is assumed that the permeability μ of the metal to which these expressions apply is equal to unity.

On differentiating (5) with respect to the temperature T , we obtain

$$\frac{dR}{dT} = \frac{4}{D^2} \left[(\epsilon - 1) \frac{dn}{dT} + 2 \frac{\sigma}{\nu} \frac{dk}{dT} \right], \quad \dots\dots(6)$$

where $D = (n+1)^2 + k^2$.

Multiplying and dividing the above expression by n :

$$\frac{dR}{dT} = \frac{4}{nD^2} \left[(\epsilon - 1)n \frac{dn}{dT} + 2n^2 \frac{dk}{dT} \right]. \quad \dots\dots(7)$$

The following should be borne in mind in the continuation of the calculation. According to Price, the region connected with the X-point is a transition region, i.e. it cannot be assumed that the wavelength under consideration is either "very long" or "very short". It is, therefore, imperative not to carry out any premature approximations, however complicated the calculation may become. The treatment below is based on the assumption that only free electrons are involved in the process of reflection: the effect of the bound electrons is to be considered later.

It has been shown by Zener (1933) that the equations (3) and (4) can be made to account for experimental results in the infra-red and visible parts of the spectrum,

if, in addition to the direct-current conductivity σ_0 , the time of relaxation τ is taken into consideration. He obtained these expressions:

$$\eta^2 - k^2 = \epsilon = 1 - \frac{4\pi N e^2}{m} \cdot \frac{1}{\omega^2 + 1/\tau^2}, \quad (8)$$

$$2nk = 2 \frac{\sigma(\nu)}{\omega} = \frac{4\pi N e^2}{m} \cdot \frac{1}{\omega\tau} \cdot \frac{1}{\omega^2 + 1/\tau^2}. \quad (9)$$

Here

$$\frac{\sigma_0 m}{N e^2},$$

where N is the number of free electrons per c.c., and e and m represent the electronic charge and mass respectively. $\sigma(\nu)$ is the conductivity corresponding to the frequency $\nu (= \omega/2\pi)$. It will simplify the presentation of the formulae if we

put $\frac{4\pi N e^2}{m} = \alpha$ and

From (3) and (4) it follows

$$\begin{aligned} 2n^2 &= \sqrt{\epsilon^2 + 4 \frac{\sigma^2}{\nu^2}} + \epsilon, \\ 2k^2 &= \sqrt{\epsilon^2 + 4 \frac{\sigma^2}{\nu^2}} - \epsilon. \end{aligned} \quad (10)$$

Substituting (8) and (9) in (10)

$$2 \frac{n^2}{k^2} = \sqrt{1 + \frac{a^2 - 2a\omega'}{\omega^2(\omega^2 + \nu_\tau^2)}} \pm 1 \mp \frac{a}{\omega^2 + \nu_\tau^2}. \quad (11)$$

Assuming that N and thus a are constants,

$$\begin{aligned} n \frac{dn}{dT} &= \left[\frac{2\omega^2 - a}{4\omega^2 \sqrt{1 + \frac{a^2 - 2a\omega^2}{\omega^2(\omega^2 + \nu_\tau^2)}}} + 1 \right] \left[\frac{a\nu_\tau}{(\omega^2 + \nu_\tau^2)^2} \cdot \frac{d\nu_\tau}{dT} \right], \\ \text{and} \quad k \frac{dk}{dT} &= \left[\frac{2\omega^2 - a}{4\omega^2 \sqrt{1 + \frac{a^2 - 2a\omega^2}{\omega^2(\omega^2 + \nu_\tau^2)}}} - 1 \right] \left[\frac{a\nu_\tau}{(\omega^2 + \nu_\tau^2)^2} \cdot \frac{d\nu_\tau}{dT} \right]. \end{aligned} \quad \dots\dots (12)$$

Substituting (8), (11), and (12) in (7), and considering only the expression which can make dR/dT equal to zero, then, if $\sqrt{\quad}$ represents the square root in (12),

$$\frac{dR}{dT} \propto \left\{ \left(-\frac{a}{\omega^2 + \nu_\tau^2} \right) \left[\frac{2\omega^2 - a}{4\omega^2 \sqrt{\quad}} + 1 \right] + \left[1 - \frac{a}{\omega^2 + \nu_\tau^2} + \sqrt{\quad} \right] \left[\frac{2\omega^2 - a}{4\omega^2 \sqrt{\quad}} - 1 \right] \right\}. \quad (13)$$

At the X-point,

$$\left[1 - \frac{a}{\omega^2 + \nu_\tau^2} + \sqrt{\quad} \right] \left[\frac{2\omega^2 - a}{4\omega^2 \sqrt{\quad}} - 1 \right] = \frac{a}{\omega^2 + \nu_\tau^2} \left[\frac{2\omega^2 - a}{4\omega^2 \sqrt{\quad}} + 1 \right].$$

A simple transformation leads to

$$\sqrt{1 - \frac{2\omega^2 - a}{\omega^2 + \nu_\tau^2} \cdot \frac{a}{\omega^2}} = \frac{2a}{2\omega^2 + a} \left(\frac{2\omega^2 - a}{\omega^2 + \nu_\tau^2} \right) - 1. \quad \dots\dots (14)$$

Writing $\omega^2 = x$, $\nu_r^2 = y$, $(2\omega^2 - a)/(\omega^2 + \nu_r^2) = z$, and squaring (14)

$$\frac{4ax}{(2x+a)^2} - \frac{4}{(2x+a)} + \frac{1}{x} = 0.$$

$$\therefore 4ax \frac{2x-a}{x+y} - 4x^2 + a^2 = 0,$$

$$x^3 - (2a-y)x^2 + \frac{3a^2}{4}x - \frac{a^2y}{4} = 0. \quad \dots\dots (15)$$

When solved, this equation gives rise to a complicated expression. But a few interesting results can be obtained without solving it.

By differentiating (15) with respect to x it is found that a maximum and a minimum occur at imaginary points. Further, when $x = -\infty$ the above expression is equal to $-\infty$; when $x=0$, the expression equals $-a^2y/4$; and when $x = \infty$ the expression is equal to $+\infty$. It follows that the above function cuts the x -axis only at one point, giving a positive value for x . Since x represents the square of ω , and the negative value of $\omega = \pm \sqrt{x}$ is rejected, it is seen that there is only one such X-point due to the free electrons.

An inspection of (13) indicates that when ω is very small as compared with ν_r and a , dR/dT will have a negative sign, as would be expected from the Hagen-Rubens relation.

It follows from (12) that the X-point is a relaxation phenomenon since (15) will be true also for $dR/d\nu = 0$.

On substituting numerical values it is found that the first two terms are negligible as compared with the latter two (15). Then, since

$$\sqrt{y} = \nu_r = 1/\tau = 2\pi c/\lambda_r$$

$$\lambda_x = \sqrt{3} \lambda_r. \quad \dots\dots (16)$$

This is in disagreement with figure 2a of Price's paper: there it would appear that λ_x was smaller than λ_r . Moreover, since λ_r is a function of the temperature, λ_x must be one likewise. The temperature coefficient of λ_r is given by that of the conductivity (Mott and Jones 1936). Therefore λ_x will be a constant for large temperature ranges only if that coefficient is negligibly small.

It will be clear from figure 1 how Price arrived at the belief that there existed an X-point whose value would be independent of the temperature. The group of curves gives a picture of the above analysis. In his work Price obtained the X-point by averaging emissivity measurements at high and reflectivity measurements at low temperatures respectively. From figure 1 it is seen that the reflectivity curves for different temperatures T_1 and T_4 cut each other at one point, but those for T_1 and T_2 do so at another, etc. It is significant that Hurst (1933) and Reid (1941), who both obtained X-points in connection with nickel, should obtain different values for this point which is supposedly characteristic of each metal, the difference being explicable in terms of the different temperature ranges in which the two workers carried out their experiments. It is only fair to admit the above theory has not tallied with all the available experimental results but to this point we shall return below. However, in spite of the scantiness of the data, it appears that bad conductors exhibit shorter wavelengths λ_x than good ones as would be expected on the basis of (16).

Theoretically speaking, the measurement of the reflectivity or emissivity, in a direction perpendicular to the surface in question, at different temperatures, and in this transition region, can provide values for τ and a ($= 4\pi N e^2/m$), which have, so far, been obtainable from catoptric measurements only. It should be pointed out that it is unlikely that the accuracy of an experimental arrangement based on this principle could emulate that of a catoptric method.

(b) *Bound electrons*

Owing to the larger number of variables in the case of bound electrons it does not seem possible to make any reasonable predictions about the temperature effect.

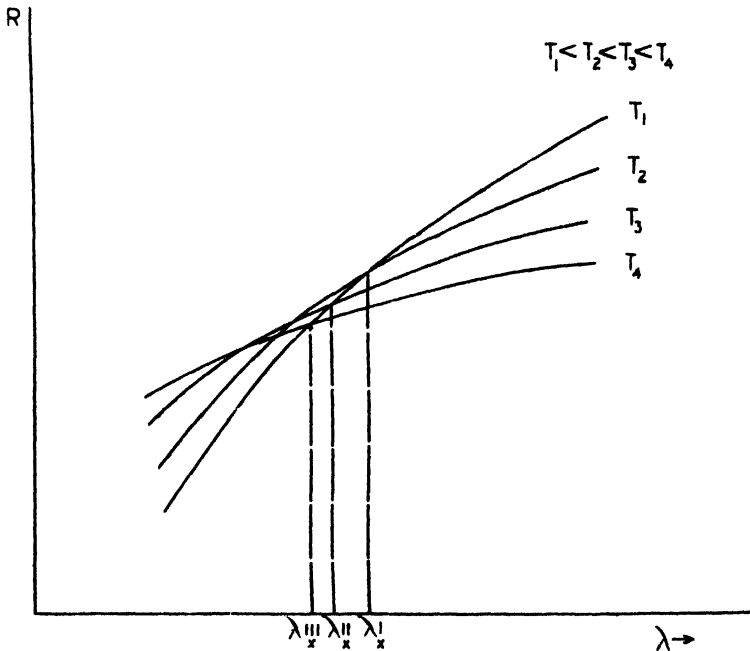


Figure 1

These variables are: the wavelength λ_0 of the absorption band, the band-width λ' , as defined by Wood (1936), and the number of the bound electrons taking part in the vibration. As might be expected the (highly complex) expression shows that R will be affected differently according as the wavelength λ is greater or less than λ_0 . Not even far-reaching approximations yield any more detailed information.

In this connection it is doubtful whether figure 3 of Price's paper is justifiable. While the absorption curves for T_1 and T_2 very likely cut each other as indicated there, it is difficult to see why the interpolated curve should cut them in the same point. Sélincourt's results for silver (1925) show a region of small temperature variation, but it is found to be on the short-wave side of the absorption band. The presence of bound electrons with a natural wavelength λ_0 in the neighbourhood of λ_x as due to free electrons only, will modify the value of the latter.

§ 3. CONCLUSION

We doubt whether an X-point as previously defined has any theoretical significance; until a more satisfactory theory is propounded it is to be assumed

that, whenever it occurs over a large range of temperatures, it is due to a fortuitous combination of the effects of free and bound electrons. However, while λ_z as defined in the present paper depends on the temperature, its possible significance in connection with optical pyrometry should not be overlooked, and will be dealt with on another occasion.

§ 4. ACKNOWLEDGMENT

The writer would like to thank Dr. H. Lowery, F.Inst.P., for his interest in the above problem.

REFERENCES

- HURST, 1933, *Proc. Roy. Soc.*, **142**, 466.
MOTT and JONES, 1926, *Theory of the Properties of Metals and Alloys*, p. 119 (Oxford : Clarendon Press).
PRICE, 1947, *Proc. Phys. Soc.*, **59**, 131.
REID, 1941, *Phys. Rev.*, **60**, 161.
SÉLINCOURT, 1925, *Proc. Roy. Soc.*, **107**, 247.
WOOD, 1936, *Physical Optics*, p. 487 (New York : The Macmillan Company).
ZENFR, 1933, *Nature, Lond.*, **132**, 968.

Work Function and Energy Levels in Insulators

BY D. A. WRIGHT

(Communication from the Research Staff of the M.O.Valve Co. Ltd.
at the G.E.C. Research Laboratories, Wembley, England)

MS. received March 1947; in revised form 2 May 1947

ABSTRACT. An estimate is made for several insulators of the energies of the highest filled energy band and of the empty conduction band. For BaO, SrO, CaO, MgO and BeO, the bottom of the conduction band is near the zero level. It is considerably lower in AgBr, ZnO and ZnS. The bearing of the results on thermionic emission, secondary emission and photoconductivity is briefly discussed, with special reference to BaO and SrO.

§ 1. WORK FUNCTIONS AND ENERGY LEVELS OF INSULATORS

CALCULATIONS concerning the position of the conduction band and of energy levels which determine the ultra-violet absorption bands were made by Mott (1938) for the case of some of the alkali halides. In these calculations it was necessary to use estimates of polarisation energies as made in a previous paper (Mott and Littleton 1938). The theory is also discussed in *Electronic Processes in Ionic Crystals* (Mott and Gurney 1940) to which frequent reference will be made. In the present paper considerations of this type are used to estimate the probable maximum value of the work function for electron insertion for some silver halides and divalent oxides. Here it is necessary to consider band widths and the effect of the selection rules.

§ 2. METHOD FOR ALKALI HALIDES

We first outline Mott's procedure in the case of the alkali halides.

The potential energy of an electron at a great distance from the crystal is indicated by the zero level O in figure 1. A is the bottom of the conduction band, so that $-\chi$ is the work necessary to introduce an extra electron into the crystal, placing it on a positive ion. C is the top of the first fully occupied band, ϕ is therefore the work to extract an electron from this band, which is usually associated with the negative ion. B is the first excited level corresponding to the formation of an exciton when an electron is removed from the full band, but remains in the field of the positive hole. If an electron is removed from this field the case corresponds with ionisation, the electron enters the conduction band, and the necessary energy is $\phi - \chi$. This energy should correspond with $h\nu_\infty$ where ν_∞ is the frequency at the series limit of the absorption bands. The energy θ corresponds with $h\nu_1$ where ν_1 is the frequency of the longest wavelength absorption band. In this treatment the width of the energy level bands was neglected, so that forbidden transitions were not considered.

The procedure was to suppose that an ion is removed from the crystal, is converted to an atom, and the atom is replaced. This either adds an electron to a positive ion or removes one from a negative ion. The method is fully discussed in Mott and Gurney (1940), pages 71, 80 and 97.

The results are

$$-\chi = W_L - I + \omega_2 + p, \quad \dots\dots(1)$$

$$\phi = W_L + E + \omega, \quad \dots\dots(2)$$

$$\theta = (2\alpha - 1) \frac{e^2}{r} + E - I + \omega_1 + p'. \quad \dots\dots(3)$$

Here W_L is the lattice energy per ion pair $= \alpha e^2/r - R$.

α is the Madelung constant for the crystal, r the interionic distance, e the electronic charge and R the energy of repulsion between an ion and its six nearest neighbours.

I is the ionisation potential of the alkali atom.

E is the electron affinity of the halogen atom.

The terms ω , ω_1 and ω_2 are polarisation energies, which will be discussed below. p is a term to allow for the fact that energy is required to insert an atom in the lattice compared with the smaller positive ion which was extracted. Such a term is omitted in (2) since an atom replaces a larger negative-ion. In (3), p' will be similar to p , but not identical with it, since in the case (3) of electron transfer, a neighbouring ion has been converted to an atom.

In these equations all the terms can be calculated except p and p' . For the alkali halides, experimental values are available for $h\nu_1 = \theta$ and for $h\nu_\infty = \phi - \chi$. Taking experimental values of $(\phi - \chi)$ and calculated values of ϕ , Mott and Gurney (p. 97) deduced values of χ . They also showed that the calculated values of $(\theta - p')$ were about 2 ev. greater than the observed values of $h\nu_1$ (see page 99 of this book).

The lattice energy W_L should, however, be used in all three cycles, whereas Mott employed it only in the first two cases. Use of the energy $W_L = \alpha e^2/r - R$ in the third cycle would give

$$\theta = (2\alpha - 1)e^2/r - 2R + E - I + \omega_1 + p'. \quad \dots\dots(+)$$

Then in all cases the value of $(\theta - p')$ would be in good agreement with the observed values of $h\nu_1$, indicating that p' is nearly zero.

It is not, however, important to discuss values of p' since the calculation of θ by equations (3) or (4) is not satisfactory. In the excited state the electron which is removed from the full band is not localised on any particular neighbouring ion. Thus the agreement with experiment is probably fortuitous, and calculations of θ will not be relied upon as an essential part of our later argument. It is perhaps preferable to regard the excited electron in the field of the positive hole as a hydrogen-like system in a medium of dielectric constant K_0 . Compared with the hydrogen case, force between charges is then multiplied by $1/K_0$, and energy separations are multiplied by $1/K_0^2$. Thus if the separation $(h\nu_e - h\nu_1)$ is known

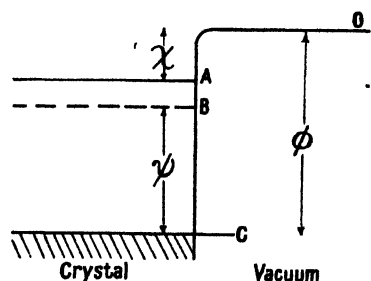


Figure 1. Energy levels in insulator.

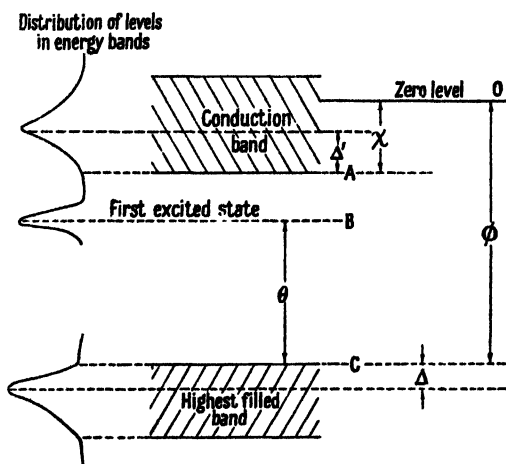


Figure 2. Energy levels in insulator including band width.

A is the bottom of the empty band.
B is the first excitation level.
C is the top of the upper fully occupied band.

for one crystal, it can be estimated for others of known K_0 since it is proportional to $1/K_0^2$. K_0 is here the dielectric constant at infra-red frequencies.

§ 3. ENERGY OF POLARISATION

The dipole formed by the electron transfer from negative ion to positive in process (3) polarises the surrounding medium, with an energy given by de Boer (1935) as

$$\omega_1 = -2.03 \frac{e^2}{r^4} (\alpha_1 + \alpha_2),$$

where α_1 and α_2 are the polarisabilities of the ions.

When an electronic charge is added or subtracted at a lattice site, the surrounding medium is polarised, and the induced dipoles produce a potential V at the site in question. The energy of this polarisation is then $-\frac{1}{2}(eV)$. Mott and Littleton (1938) gave methods for calculating ω and ω_2 , which are also described in Mott and Gurney (1940) pages 58 and 59. Our procedure has been to use Mott's approximation of the first order, and to calculate the potentials when an electron is

added at a positive ion site, ω_2 , or subtracted at a negative ion site, ω . In all cases the estimation of α_1 and α_2 has been made taking the observed value of K_0 which gives $(\alpha_1 + \alpha_2)$ according to the formula

$$\frac{K_0 - 1}{K_0 + 2} = \frac{4\pi}{3} \frac{1}{r^3} \frac{\alpha_1 + \alpha_2}{2}.$$

Then we take the observed ratio α_1/α_2 for the free ions and determine the values of α_1 and α_2 from α_1/α_2 and $\alpha_1 + \alpha_2$. The values found in this way are shown in the tables below. where observed values are those given by Mott and Gurney (1940).

§ 4. BAND WIDTH

In the considerations so far, band width has been neglected. We shall make no attempt to calculate band widths in the insulators in which we are interested, but shall consider the experimental evidence available, and the effect of observed band widths on calculations of the type discussed above. Now O'Bryan and Skinner (1940) showed that in the alkali halides the width of the highest filled band is from 5 to 19 ev., and in a variety of oxides it is from 14 to 23 ev. Such widths must have a considerable effect on our calculations. However, the shape of the bands must be considered, and we note that in most cases a large part of the width is due to ends where the electron density is very low. The main part of the band is frequently a single sharp peak which may be compared with a broadened spectrum line. In such cases the "line-width", which is the width of the band at half maximum intensity, is considerably smaller than the total width. This is especially true of the alkali halides, and among the oxides of BaO, SrO and CaO, where the line-width is only 2 ev., although the total width is 14 ev. The fact that there is some experimental agreement in the case of the alkali halides makes it doubtful whether the band ends have sufficient density to affect the optical behaviour of the crystal.

We shall suppose here that the calculations leading to equations (1) and (2) above predict the distance below the zero level of the peaks of the energy bands, as indicated in figure 2, and we shall also suppose that the band widths are the widths without the low density ends. Thus for BaO with total width 14 ev., at either side of the band there are ends extending for 4 ev., and we shall take the effective width as 6 ev. The peak is then 4 ev. below the top of the band. It should be noted that apart from the effects of low density of electrons in the top end of the filled band and of low density of levels in the lower end of the conduction band, the transition from the top of the filled band to the bottom of the conduction band is forbidden, and the frequency for the series limit will therefore in any case be greater than that which corresponds to the separation between the band ends. Similarly the transition from the top of the full band to the first excitation level is forbidden. We shall continue to calculate the energy θ from equation (4) for the sake of completeness, and because of the agreement with experiment in the alkali halides; θ should presumably correspond with the separation between the peak of the full band and the peak of the band of excitation levels. However, as pointed out above, the calculation is really unsatisfactory in this case.

We shall in the following refer to figure 2 in our calculations, thus equation (2) determines the depth of the peak of the band below the zero level, and we write ϕ

for the depth of the top of the band, and Δ for the depth of the peak below the top. Thus equation (2) determines $(\phi + \Delta)$, and equation (2) determines $-(\chi - \Delta')$. We shall continue to write (θ) for the energy determined by equation (4).

§ 5. NaCl INCLUDING BAND WIDTH

We first consider the case of NaCl when band width is not neglected. The quantities necessary for the cycle calculations are shown in table 1.

We find

$$\begin{aligned}\phi + \Delta &= 10.2, \\ -(\chi - \Delta') &= 0.3 + p, \\ \theta &= 7.7 + p'.$$

We know that the 3s band in the alkali chlorides is narrow and has short ends, thus Δ in this case is about 0.5 ev., hence $\phi = 9.7$. The excitation level, which is known from experiment to lie 7.7 ev. above C, is therefore 2 ev. below the zero level 0, and the series limit is a little higher than 0. Now we know that pure NaCl is not photo-conducting when irradiated in its first absorption band, thus the bottom of the empty conduction band is considerably higher than the first excitation level, and is therefore not more than about 1 ev. below the zero level. Thus the maximum value of χ is about 1 ev., which is consistent with Mott and Gurney's experimental value of 0.5 ev., page 74. If p is nearly zero, this indicates that Δ' is about 1 ev. Note that when band widths are included the position of the bottom of the conduction band is not necessarily determined by adding to the calculated value of ϕ the observed value of $h\nu_{\alpha}$, as it was when band widths were neglected. There may now be forbidden transitions with energy lower than $h\nu_{\alpha}$, so that χ may be greater than zero, and the argument from photo-conductivity becomes necessary.

Table 1

	NaCl	AgBr	AgCl
α	1.75	1.75	1.75
r	2.81	2.88	2.77
ae^2/r	8.9	8.6	9.0
K_0	2.33	4.62	4.01
α_1 for halide ion	3.0	3.0	3.0
α_2 for metal ion	0.15	4.1	2.2
ω_1	-1.6	-2.7	-2.7
ω_2	-2.5	2.9	-3.2
ω	-1.5	2.8	-2.6
I	5.12	7.54	7.54
E	3.75	3.55	3.75

§ 6. SILVER HALIDES

For AgBr, using the figures in table 1, we find:—

$$\begin{aligned}\phi + \Delta &= 8.5, \\ -(\chi - \Delta') &= p - 3.1, \\ \theta &= 3.5 + p'.$$

For AgBr, O'Bryan and Skinner found that the width of the main part of the 4p band is 4 ev. (1940), and the peak lies 1 ev. below the top of the main part, consequently $\Delta = 1$ ev. There is, however, a considerable tail at the top of the band.

The optical absorption evidence (Fesefeldt and Gyulai 1929) shows a rapid increase in absorption beginning near 3000 Å., which indicates that the excitation level B lies about 4 ev. above C (figure 2). The high value of K_0 shows that the series limit energy, $h\nu_{\infty}$, will only be a little higher than $h\nu_1$, and, in fact, since it is known that photo-conductivity occurs in the tail of the absorption band, the bottom of the conduction band must almost coincide with B. This gives $\chi \simeq 8.5 - 1 - 4 = 3.5$ ev. This is in good agreement with Mott's value of 3.6 deduced from experiments on the bleaching of emulsions (Mott and Gurney, p. 245). It will be noted that as with the alkali halides, θ agrees roughly with the observed excitation energy if p' is small, and the value of $\chi - \Delta'$ is similar to the value deduced for χ if p is small, indicating that Δ' is probably less than 1 ev. The long optical absorption tail extending into the visible may correspond with transitions from the high-energy tail in the 4p band.

For AgCl we find: —

$$\begin{aligned}\phi + \Delta &= 9.0, \\ -(\chi - \Delta') &= p - 2.4, \\ \theta &= 4.2 + p' .\end{aligned}$$

There is no information concerning the width of the 3s band in AgCl, consequently we do not know Δ . As with AgBr, the optical absorption increases rapidly near 4 ev. (Fesefeldt and Gyulai 1929), and Mott here also deduces that χ is about 3.6. This would be consistent on the above argument with $\Delta = 1.5$ ev., and with Δ' a little greater than 1 ev.

§ 7. DIVALENT OXIDES

We suppose that the oxides we shall discuss are in the form of ionic crystals and that the ions are both doubly charged in the normal state.

Now, when the doubly charged positive ion is removed from the crystal, the energy required is $4\alpha e^2/r - R$, and addition of an electron requires the energy $-I_2$, where I_2 is the ionisation potential for the second electron. Replacement in the crystal involves the energy $2\alpha e^2/r - R'$, where R' is the energy of repulsion between the singly charged ion and its nearest neighbours. Here we shall neglect the difference between R and R' . We thus obtain

$$-(\chi - \Delta') = 2\alpha e^2/r - I_2 + \omega_2 + p. \quad \dots\dots(5)$$

Similarly,
$$\phi + \Delta = 2\alpha e^2/r + E_2 + \omega. \quad \dots\dots(6)$$

In the case of the transfer of an electron from a negative to a positive ion, corresponding with excitation, the cycle gives

$$\theta = (4\alpha - 1)e^2/r + E_2 - I_2 + \omega_1 + p'. \quad \dots\dots(7)$$

In these equations, as before, p and p' cannot be satisfactorily calculated. Experimental values of Δ are available, but there is no reliable information about Δ' . Some approximate estimates will be made below. The other quantities are known accurately, except that there may be some uncertainty about the value of E_2 , the electron affinity of the singly-charged negative oxygen ion. We take the value deduced by Seitz (1940) of -9 ev., though this might require some modification in

the future. This was obtained by first estimating the affinity of oxygen for two electrons from the results of Sherman (1932) and of Mayer and Maltbie (1932), and then noting that the affinity of neutral oxygen for one electron is +2 ev. (Lozier (1934)).

We are assuming in the above equations that the crystals are completely polar. This was sufficiently true in the case of the alkali halides, and the results of O'Bryan and Skinner (1940) indicate that it is probably a satisfactory assumption for the present purpose in the case of BaO and SrO. They concluded, however, that MgO and BeO are much less completely polar, so that our results will be less accurate in these cases.

The quantities necessary in the calculations are shown in table 2. Here the polarisation energies have been obtained accurately for the cubic crystals, taking Mott's first-order approximation as with the halides. In the case of BeO and ZnO, with a wurtzite lattice, the necessary summations are not available, and we have assumed that for the observed values of the interionic spacing r and the dielectric constant K_0 , the terms are similar to those for corresponding values in a cubic lattice.

Table 2

	BaO	SrO	CaO	MgO	BeO	ZnO	ZnS
a	1.75	1.75	1.75	1.75	1.64	1.64	1.64
r	2.77	2.57	2.40	2.10	1.64	1.94	2.33
ae^2/r	9.0	9.8	10.4	11.8	14.4	12.0	10.1
K_0	3.5	3.31	3.28	2.95	2.95	—	5.07
a_1 for negative ion	3.0	2.65	2.9	1.64	0.85	—	3.36
a_2 for metal ion	1.6	0.89	0.6	0.05	0	—	0.12
ω_1	— 2.3	— 2.4	— 2.4	— 2.5	— 3.5	— 2.5*	— 3.5
ω_2	— 2.8	— 2.8	— 2.9	— 3.0	— 2.9	— 3.0*	— 3.7
ω	— 2.4	2.3	2.0	— 1.7	— 1.6	— 1.5*	— 2.0
I_2	10.0	11.0	11.8	15.0	18.1	17.9	17.9
E_c	— 9.0	— 9.0	— 9.0	9.0	9.0	9.0	— 5.0

* These values are assumed by comparison with the other compounds since K_0 is not known.

§ 8. BARIUM AND STRONTIUM OXIDES

In table 3 we show the results of the application of equations (5), (6) and (7), the distance Δ from the top of the main part of the filled band to the peak as determined by O'Bryan and Skinner, and hence the value of ϕ .

Table 3

	BaO	SrO
$-(\chi \quad \Delta')$	5.2 + p	5.8 + p
$\phi \quad \Delta$	6.6	8.3
θ	9.7 - p'	11.2 + p'
Δ	4.0	5.0
ϕ	2.6	3.3

There is no direct optical information in either case concerning the values of $h\nu_1$ and $h\nu_g$, however, since the crystals do not absorb in the visible, $h\nu_1$ is at least 3 ev. and is probably considerably greater. Thus both the first allowed excitation level and the first ionisation level must lie above the zero line. The bottom of the conduction band will be lower than the first allowed ionisation level, and may in

fact be below the zero line. In the case of BaO we know it cannot be much below the zero line however, since impurity levels can be added to BaO making it a semiconductor. When excess Ba is present, the crystal becomes semi-conducting with an activation energy varying from 0.5 to 1 ev. according to the experimental conditions. The distance below A (figure 2) of the impurity levels may therefore be as large as 2 ev.; thus C must lie at least 2 ev. below A. Since ϕ is 2.6, we deduce that the maximum value of χ is 0.6 ev. It may of course be less, and will be negative if the bottom of the conduction band is above the zero level.

The experimental results having a bearing on these conclusions will be more fully discussed in the following paper. We may note here, however, that Nishibori and Kawamura (1940) found for both SrO and the equimolecular double oxide (BaSr)O, a value of χ of 0.3 ev., and that they concluded that these values of χ were typical of the oxides themselves, and were not affected by conditions at the outer surface of the oxides, e.g. by the presence of adsorbed Ba. It was primarily in order to find evidence concerning this conclusion that the present calculations were carried out. It will be noted that if we take this experimental value for χ and assume p to be small, as in the case of the halides, we deduce that Δ' is 5.5 ev. for BaO and 6.1 ev. for SrO. These are the depths of the bottom of the conduction levels below their peak. It can also be deduced from the results of Nishibori and Kawamura that the photoelectric work function of the unactivated oxides lies between 2.8 and 3.7 ev. for SrO, and between 2.0 and 2.7 ev. for (BaSr)O. These correspond well with our values for ϕ , as they should if we assume there is no free impurity barium in the unactivated crystal in these experiments.

§ 9 OTHER DIVALENT COMPOUNDS

In table 4 we show the results on some further oxides and on zinc sulphide, where following Seitz (1940) we take the value of E_g as -5 ev.

Table 4

	CaO	MgO	BeO	ZnO	ZnS
$(\chi \quad \Delta')$	$6.1 + p$	$5.6 + p$	$8.2 + p$	$3.0 + p$	$-1.3 + p$
$\phi \quad \Delta$	9.8	12.9	18.2	13.5	13.2
θ	$12.5 + p'$	$14.0 + p'$	$18.2 + p'$	$11.5 + p'$	$8.3 + p'$
Δ	5.0	7.0	7.0	4.0	3.0
ϕ	4.8	5.9	11.2	9.5	10.2

There is little experimental evidence concerning these compounds. For MgO and BeO we have the indirect information obtained by O'Bryan and Skinner, and discussed in Mott and Gurney (1940), pages 77 and 78, which indicates that $h\nu_1$ lies between 11 and 15 ev. for MgO, and is near 14 ev. for BeO. Here again, therefore, excitation raises the electron above the zero level. If we identify θ with $h\nu_1$, we deduce that p' in these cases lies between -2 and -4 ev. We can obtain approximate values of Δ' in the case of BeO and MgO by Seitz's method (1938). In the case of ZnS, Seitz compared the broadening of the 4s level of Zn^+ with the broadening of the comparable 4s level in Cu, which was studied by Fuchs, and deduced that the depression corresponding with our Δ' was 4 ev. Similarly we can compare the 2s levels of Be^+ in BeO with those of metallic Li, and the 3s levels of Mg^+ with those of metallic Na, both of which metals have

been studied by O'Bryan and Skinner (1934). Since the separation between the metal ions in the oxides is smaller than between Li and Na atoms in the metal, we conclude that approximately for Be^+ , $\Delta' = 5$ ev. and for Mg^+ , $\Delta' = 4$ ev. Then if we suppose that p is of the same order as p' , we find that for both MgO and BaO , χ is nearly zero. We cannot of course predict a definite value, but it is unlikely that χ is greater than 1 ev., and it could have a small negative value.

In the case of ZnO , it is fairly certain that χ has a positive value, and the probable limits are from +1 to +3 ev. It is doubtful whether the excitation level is as high as the zero line, but it is certainly higher than the bottom of the conduction band. Our conclusions differ from those of Seitz (1940), since he did not include either the polarisation terms or the width of the negative ion band. This applies equally to the case of ZnS , which Seitz considered in 1938 and 1940. Here χ is of the order +5 ev. In the case of ZnS , as in all the compounds dealt with in this section, the separation between the bands is considerably smaller than the first excitation energy, so that we expect photoconductivity in the tail of the first optical absorption band. This is observed in ZnS .

§ 10. THERMIONIC EMISSION

We have found that for all the oxides in §§8 and 9 except ZnO , the predicted value of χ is nearly zero, and that the maximum value which appears possible increases from 0.5 to 1.0 ev. in the order Ba Sr Ca Be Mg, though in each case the value might be negative. There is some experimental evidence to show that the value is 0.3 for BaO and SrO . Now, when a metal impurity atom is added with an electron energy-level at a depth 2ϵ below the bottom of the conduction band, the oxide becomes a semi-conductor with activation energy of conductivity ϵ . The thermionic work function is nearly $\chi + \epsilon$. We expect the value of ϵ to be related to the ionisation potential of the metal atom, and therefore to be larger, for example, for Be than for Ba. Thus if the value of χ were the same for BeO as for BaO , we should nevertheless expect the thermionic work function of BeO containing Be to be greater than that of BaO containing Ba, as is observed. If, however, Ba could be introduced into one of the other oxides whose χ was similar to that of BaO , it might be possible to obtain the same thermionic work function as that of BaO , and of course if one of the other oxides had a smaller χ than that of BaO , a correspondingly smaller work function might be obtained. Intermediate values would be expected with a mixture of oxides in the case where they form a solid solution. These considerations apply to the mixtures of BaO and SrO , and of BaO , SrO and CaO which are commonly employed as coatings for oxide cathodes. In such mixtures, solid solutions are formed, and when free metal is produced by a reduction or electrolytic process, the resultant metal is mainly Ba, since the heat of formation of BaO is less than that of the other oxides. It is known that the work function of such coatings is of the same order as that of BaO , and may be slightly less, although the work functions of SrO containing Sr and CaO containing Ca are considerably greater. This is understandable according to the above argument. We therefore deduce that the values of χ for SrO and CaO are similar to the value for BaO . An accurate comparison will be possible when the work functions of the mixed oxides have been measured accurately over a range of composition. Mixtures of BaO with MgO and BeO do not appear to have been studied.

§ 11. SECONDARY EMISSION

It is well known that all the oxides considered except ZnO have high secondary emission. Bruining and de Boer (1939) suggested that in the insulators with high secondary emission, the first empty band is situated above the zero level, so that when the incident primary electron causes a transition, the secondary raised from the full negative ion band can readily escape. The present results indicate that though the whole of the conduction band is probably not situated above the zero level, the first allowed transition always raises the electron above the zero level, which is in agreement with the general argument of Bruining and de Boer's theory. The selection rules for transitions caused by electron bombardment are similar for primary energies of several hundred volts to those for optical transitions, according to Wooldridge (1939). There does not appear to be any information concerning ZnO, but since χ is probably nearly 3 ev., and the excitation level is near zero, it is not likely to be as good a secondary emitter as the other oxides.

REFERENCES

- DE BOER, 1935, *Electron Emission and Absorption Phenomena* (Cambridge), p. 241.
 BRUINING and DE BOER, 1939, *Physica*, **6**, 834.
 FESEFELDT and GYULAI, 1929, *Nachr. Wiss. Göttingen*, 229.
 LOZIER, 1934, *Phys. Rev.*, **46**, 268.
 MAYER and MALTBIE, 1932, *Z. Phys.*, **75**, 748.
 MOTT, 1938, *Trans. Faraday Soc.*, **34**, 500.
 MOTT and GURNEY, 1940, *Electronic Processes in Ionic Crystals* (Oxford).
 MOTT and LITTLETON, 1938, *Trans. Faraday Soc.*, **34**, 485.
 NISHIBORI and KAWAMURA, 1940, *Phys. Math. Soc. Japan*, **22**, 378.
 O'BRYAN and SKINNER, 1934, *Phys. Rev.*, **45**, 370.
 O'BRYAN and SKINNER, 1940, *Proc. Roy. Soc., A*, **176**, 229.
 SEITZ, 1938, *J. Chem. Phys.*, **6**, 454.
 SEITZ, 1940, *Modern Theory of Solids* (McGraw-Hill), p. 448.
 SHERMAN, 1932, *Chem. Rev.*, **11**, 93.
 WOOLDRIDGE, 1939, *Phys. Rev.*, **56**, 562.

Energy Levels in Oxide Cathode Coatings

By D. A. WRIGHT

(Communication from the Research Staff of the M.O. Valve Co. Ltd.
 at the G.E.C. Research Laboratories, Wembley, England)

MS. received 1 March 1947; in revised form 21 May 1947

ABSTRACT. The energy level diagram is considered in the case of a Ba/SrO emissive cathode coating, and it is concluded from the results the preceding paper that the process of activation consists in building up a concentration of free barium in the coating, and in providing barium at the interface between core and coating. The work function of the activated coating without adsorbed barium on its outer surface cannot be much greater than the observed value of 1 ev., and may be no greater, so that such an adsorbed layer, if present, has only a small effect on the work function. The possibility of increase in emission under the influence of ultra-violet light or electron bombardment is briefly discussed.

§ 1. THE METAL-COATING CONTACT AND THE WORK FUNCTION

THE oxides discussed in the preceding paper have not been investigated directly as regards energy levels, but indirect information is available from other studies. In the present paper we are concerned with BaO and SrO, which in the mixed oxide form are employed as thermionic emitters. The double oxide Ba/SrO is applied to a metallic base, and in the highly emitting state is a semi-conductor activated by the presence of excess barium. In studying the mechanism of emission, measurements have been made of the thermionic work function, the electrical conductivity, the photoelectric behaviour and the secondary emission. There has been considerable variation in results obtained, as a result of which the details of the mechanism remain uncertain.

Three types of picture have been proposed: (a) According to de Boer (1935) barium atoms are adsorbed on the outer surface of the coating, and can lose their electrons by thermal ionisation. The ionisation energy of the adsorbed atom is less than that of a free atom. The emitted electrons are replaced by a flow of electrons through the coating. (b) According to Reimann (1934) the electrons are emitted from the conduction band of the semi-conductor. It is supposed that an adsorbed layer of barium reduces the effective work function compared with that of the "clean" coating, precisely as when such a layer is adsorbed on the surface of a more electro-negative metal. (c) It is possible that the emission is the inherent emission of the coating as in (b), but that the work function of the "clean" coating is sufficiently low to lead to high emission. In this case it would be possible for high emission to be obtained in the absence of an adsorbed layer of barium. In all three theories it is necessary to suppose that excess barium is freed in the coating, thus raising its conductivity compared with that of the pure oxide. This barium is freed by electrolysis or by chemical reaction with reducing agents in the core metal or coating. At the interface between core and coating a potential barrier is to be expected, whose height we shall assume for the moment to be equal to the difference between the work functions of core metal and coating. The barrier may be of the Schottky type, or a layer may be present between core and coating whose resistance is higher than that of the coating. These barriers have been discussed by the writer (1947), and evidence for their presence has been given by Fineman and Eisenstein (1946).

The thermionic work function of the coating is the height OE in figure 1, where O is the potential energy at infinity, A is the bottom of the conduction band in the crystal, and C the top of the full band as in the preceding paper. D is now an added level due to the Ba impurity, and the level E is approximately half-way between A and D.

When the coating is in equilibrium with the core metal, the levels in the metal are raised until the top of the Fermi level in the metal F is at the same height as E, figure 2. Thus PF is the difference between the two work functions. Now experiments by the writer have indicated that the value of PF is 0.7 ev. when the core metal is nickel containing a small proportion of magnesium. The work function of Ni is 5.0, thus if the magnesium is unimportant in influencing the work function, the coating work function must be 4.3. Now there is good experimental evidence that the thermionic work function of the activated coating is of the order 1 ev. This appears at first to rule out picture (c) above, and to require either (a)

in which the energy levels are not directly involved, or (b) in which an adsorbed layer of barium on the outer surface reduces the effective work function. The energy level system is then as in figure 3.

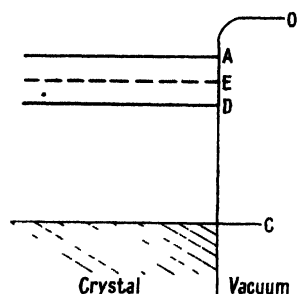


Figure 1. Energy levels in excess semi-conductor

C is the top of the upper fully occupied band

A is the bottom of the conduction band.

D is the added impurity level

E is half way between A and D.

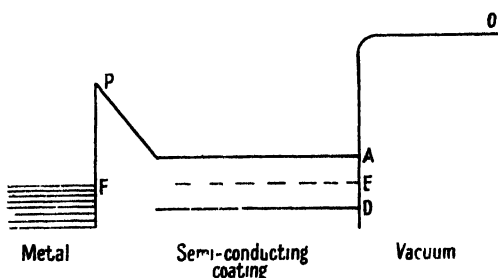


Figure 2. Excess semi-conductor in contact with metal. Case where semi-conductor surface is "clean"

A, E and D are levels as in figure 1

F is the top of the Fermi levels in the metal

PF is the difference in work functions of metal and semi-conductor

This type of diagram was discussed by Reimann (1934), except that he considered the barrier at the metal-coating interface to be thin enough for tunnel effect penetration. The writer has indicated elsewhere (Wright, 1947) that in many cases the barrier is too thick for this to occur. Here the energy level diagram discussed was as in figure 3, i.e. it was supposed that the adsorbed barium layer might be present at the outer surface.

There were, however, clear indications that the behaviour of the coating was determined more by the conditions at the metal-coating interface than by the conditions at the outer coating surface. Thus the time-decay effects in thermionic emission current and in conduction current were related and were both associated with the core-metal boundary. Thus if mechanism (a) is applicable, we suppose that changes at the interface affect the conductivity, and through it the emission, because of the electron supplementation process. It seems necessary to attribute the decay effects to electrolytic flow, which led the writer (1947) to the tentative proposal that positive ions are adsorbed at the metal-coating interface. Alterations in their distribution then leads to the decay effects in emission and conduction current. Now it is important to note that this possibility alters our conclusion above that mechanism (c) is ruled out. Thus if we assume that the work function of the "clean" oxide without adsorbed barium has the observed value of 1.0, then with $PF=0.7$ we require the work function of the nickel with an adsorbed layer

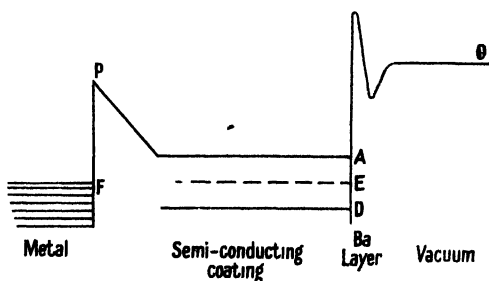


Figure 3. Excess semi-conductor in contact with metal. Case where adsorbed electro-positive layer is present on semi-conductor surface. Levels as in figure 2.

of barium to be 1.7, which is quite a possible value. Thus we have one extreme possibility consistent with (b), that the work function of clean Ba/SrO is 4.3, but an adsorbed layer of Ba on the outer surface reduces the effective value to 1.0, or the opposite extreme consistent with (c), that the work function of the coating is 1.0, but an adsorbed layer of Ba at the metal-coating interface reduces the height of PF from 4.3 to 0.7. The decay phenomena indicate that (b) is not a true picture. The alternatives are (c), a state intermediate between (b) and (c), or de Boer's mechanism (a).

It should be noted that the experimental evidence concerning the work function does not in most cases throw any light on this question. The fact is established that the thermionic work function of BaO or of Ba/SrO is nearly 1 ev., but most results give no indication whether the true situation resembles figure 2 or figure 3. In the case of the work of Nishibori and Kawamura (1940), referred to in the preceding paper, it was found that the value of χ remained at 0.3 ev. during activation, although the thermionic work function decreased by 0.4 ev. This indicated that the only effect of the freeing of Ba during the activation lay in increasing the concentration of impurity levels D and consequently raising the level of E. Thus this evidence was strongly in favour of figure 2. It is, however, possible to argue that surface Ba was present at the beginning of the activation in these experiments, and that its concentration did not vary during activation. If this were the case, the evidence would not be conclusively in favour of figure 2. It seems unlikely in fact that a surface concentration could be maintained constant, independent of the internal concentration; however, in view of this doubt, the calculations in the preceding paper were carried out. These show that the value of χ is not greater than 0.6 for BaO without a surface layer of Ba. Since the experimental values for the activation energy of conductivity AE lie between 0.5 and 1.0 ev., we conclude that the thermionic work function in the absence of adsorbed Ba cannot be greater than 1.6. This rules out the extreme case referred to above where the work function of clean oxide is 4.3 ev., and the adsorbed layer of Ba is very important in reducing the work function after activation to 1.0 ev. It permits the other extreme case corresponding with figure 2, and theory (c), in which the work function of the clean oxide can become 1.0 after activation, without the presence of adsorbed Ba. It also permits intermediate cases in which, with an adsorbed layer of Ba present, the work function is slightly lower than in the absence of such a layer, i.e. lower by a few tenths of an electron volt. This would correspond with a form of figure 3 in which the depression of the work function due to the adsorbed layer is small.

We conclude therefore that the thermionic emission from BaO or Ba/SrO is emission from the conduction band of the semi-conductor, the bottom of which is situated not more than 0.6 ev., and probably only 0.3 ev. below the zero level. The process of activation consists in building up a sufficient concentration of free Ba in the oxide to lower the work function to 1 ev. by raising the level E, and in providing a layer of adsorbed Ba at the coating-core interface. At the outer coating surface, while Ba must be present at least intermittently, since it evaporates from the oxide, it does not play a very important part in lowering the work function to the observed value of 1 ev. in the activated cathode. There is an indication that it may play no part at all, i.e. that the clean oxide with an adequate internal

concentration of free Ba has a work function of 1 ev. If the adsorbed layer of Ba does play a part, it can only be small, since the work function of the clean oxide cannot be much more than 1 ev. We may note here that the conclusions of Huber and Wagener (1942) were that there is no electron diffraction evidence to support the view that a layer of adsorbed Ba is present on an activated cathode.

§ 2. INFLUENCE OF LIGHT AND ELECTRON BOMBARDMENT

When electrons are raised into the conduction band of a semi-conductor, for example, under the influence of light of sufficiently short wavelength, the conductivity is increased. The increase in the number of electrons in the conduction band which causes this should also increase the thermionic emission from the semi-conductor. Apart from this short-time effect, there may be a photochemical decomposition, as in the alkali and silver halides, where loss of an electron from a negative ion can lead to the loss of a negative ion from the crystal. If there is a corresponding effect in the oxides, there should be an increase in conductivity and in thermionic emission from the oxide under the influence of the light, if the oxide is not initially fully activated.

There is some evidence that one or both of these processes may occur in oxide coatings (Case, 1921; Merritt, 1921; Crew, 1926; Newbury, 1929; Bodemann, 1929; Ramanadoff, 1931). In these papers it was found that there was a rise in photoelectric response in the near ultra-violet with temperature. It was at first thought that there was an increase in photo-electric emission with temperature; it later appeared, however, that the true effect was an enhancement of the thermionic emission under the influence of light of sufficiently short wavelength. There was no effect above 3000 Å., a small effect from 3000 to 2500 Å., and a considerable increase near 2500 Å. By analogy with the results on the photo-conductivity of silver halides, it can be deduced from these observations that the range 5 to 4 ev. lies in the tail of the first optical absorption band, whose peak will therefore occur at considerably higher energy. In the case of electron bombardment of an oxide, similar effects might be expected. Johnson (1946) obtained effects which he interpreted as the enhancement of emission from poorly activated oxide cathodes under electron bombardment, though the interpretation is doubtful in view of the later results of Pomerantz (1946), who found a true increase in secondary emission with temperature. As regards the photo-chemical effect, both Headrick and Lederer (1936) and Jacobs (1946) have obtained evidence for the liberation of oxygen from oxides under electron bombardment. If this does occur in oxide cathodes, it is possible that continued bombardment of an activated cathode would lead to the formation of an excess of free metal and hence of metal aggregates, though such an effect would be temperature dependent. Their formation at low temperature might explain the observed decay in secondary emission with time, since metal aggregates may decrease the secondary emission both by replacing the oxide and by providing electron traps. At higher temperatures, for example 700 or 800° C., the excess metal would evaporate, and the loss of oxygen and metal would lead to disappearance of coating at a greater rate than in the absence of electron bombardment. Such an effect is observed in cavity magnetrons where some of the emitted electrons return to bombard the cathode.

REFERENCES

- BODEMANN, 1929, *Ann. Phys., Lpz.*, **3**, 614.
 CASE, 1921, *Phys. Rev.*, **17**, 388.
 CREW, 1926, *Phys. Rev.*, **28**, 1265.
 DE BOER, 1935, *Electron Emission and Adsorption Phenomena* (Cambridge), p. 362.
 FINEMAN and EISENSTEIN, 1946, *J. Appl. Phys.*, **17**, 643.
 HEADRICK and LEDERER, 1936, *Phys. Rev.*, **50**, 1094.
 HUBER and WAGENER, 1942, *Z. tech. Phys.*, **23**, 1.
 JACOBS, 1946, *J. Appl. Phys.*, **17**, 596.
 JOHNSON, J. B., 1946, *Phys. Rev.*, **66**, 352.
 MERRITT, 1921, *Phys. Rev.*, **17**, 525.
 NEWBURY, 1929, *Phys. Rev.*, **34**, 1418.
 NISHIBORI and KAWAMURA, 1940, *Phys.-Meth. Soc. Japan*, **22**, 378.
 POMERANTZ, 1946, *Phys. Rev.*, **70**, 33.
 RAMANADOFF, 1931, *Phys. Rev.*, **37**, 884.
 REIMANN, 1934 *Thermionic Emission* (Chapman and Hall), p. 228.
 WRIGHT, D. A., 1947, *Proc. Roy. Soc., A*, **190**, 394.

The Experimental Basis of Electromagnetism: Part II—Electrostatics

BY N. R. CAMPBELL AND L. HARTSHORN

MS. received 28 March 1947

ABSTRACT. This paper continues the argument of a previous paper with the same title. It explores the foundations of the science of electrostatics as practised in the modern laboratory, admitting as evidence only those experiments that experience has proved to be practicable. The basis of the subject is found in alternating currents and the laws of capacitance, which lead to the conception of electric potential energy. The various experiments on the mechanical forces between electrified bodies fit in well with this conception, and we are led to the conclusion that the soundest procedure in the investigation of any system of conductors and dielectrics is to represent it by its equivalent capacitance network. Coulomb's Law and point charges play no part in our scheme, but they are discussed because of the prominent part that they play in classical theory. The familiar equations for point charges, which are the premises from which that theory starts, can be regarded as definitions of what is meant by "point charge", or as approximations to the laws governing the forces on small charged bodies, which become true in a limiting case. One incidental advantage of our treatment is that the somewhat elusive notion of "earth" in the usual expositions is replaced by more definite conceptions.

§ 1. ALTERNATING CURRENTS

THE general principles to be followed in this enquiry were outlined in a previous paper (1946), here called Part I, which established the magnitudes characteristic of D.C. circuits, viz. current, voltage, resistance etc. Our present object is to show that the working principles of electrostatics as practised in the modern laboratory can be soundly based on real experiments, and in this way to indicate the true nature of the experimental foundations of the subject. It is a commonplace among experimental workers that the magnitudes that appear in electrostatics are today almost invariably measured by alternating current methods.

which are far more accurate than those used by the pioneers in the subject; we naturally therefore look to alternating currents for our basic conceptions.

In one of the circuits from which we started, let the battery be replaced by an A.C. generator (say audio-frequency). Current indicators in series will not generally read the same current, according to their D.C. calibration; nor will voltage indicators in parallel read the same voltage. If the indicators are of the kind generally employed in modern practice—but not necessarily otherwise—most of the direction-sensitive indicators will read nothing at all, while the reversible indicators will continue to agree among themselves. If we had started from such an arrangement, we should simply have ignored the direction-sensitive indicators as not indicators at all.

But, even then, if we had tried to calibrate the reversible indicators by addition in the manner described in Part I, we should have failed, because the necessary laws of addition are not true in general. They are true if the circuit elements are limited to a narrow class (that of “pure” elements all of the same kind, see §2); and this fact is used in the calibration of A.C. instruments according to the best modern practice. But, though we might by such methods measure independently magnitudes corresponding to current and voltage, it would be doubtful whether we should be wise to call them by those names. For there would be no negative currents or voltages; and the laws from which “current” and “voltage” derive so much of their significance would not be true; neither the sum of the currents flowing to a node nor the sum of the voltages round a circuit would be zero.*

However, if we include cathode-ray oscillographs among the indicators, we shall find that they, though direction-sensitive, continue to give an indication; but they indicate that the state of the system is not constant, but is varying cyclically. If we calibrate these instruments by D.C. for current and voltage, and by some form of stroboscope for time, and if we then regard only those values of current and voltage that are simultaneous, we shall find that simultaneous current and voltage possess over a wide range the properties of direct current and voltage; in particular they obey the laws from which these terms derive their significance. Moreover, we shall find that the behaviour of all the instruments that give steady indications in an A.C. circuit can be completely explained in terms of their reactions to different direct currents and of the presence of some kind of inertia that causes them to read a mean value of a rapidly varying influence. Electrostatic and dynamometer instruments agree among themselves because they all read nearly the same mean; rectifier instruments diverge because they read a different mean.

These are the considerations that justify the conceptions of instantaneous and RMS current, voltage and power, and that underlie the calibration of RMS instruments in terms of D.C. in standardizing laboratories. Certain elaborate precautions have to be observed; but since they do not involve any of the laws we are about to discuss, they need not detain us.

§ 2. PURE RESISTORS AND PURE CAPACITORS

When the A.C. generator is substituted for the battery, none of the laws concerning circuit elements set forth in Part I remains generally true. In particular, if equality of impedance or admittance means equivalence in respect of a reversible indicator in an A.C. circuit, impedance is not generally additive in series

or admittance in parallel; consequently impedance and admittance are not, like resistance and conductance, measurable by addition. But it is possible by certain tests to select, from the whole class of elements that may form part of an A.C. circuit, two sub-classes that possess characteristics that are additive and are therefore measurable independently; they will be termed respectively "pure resistors" and "pure capacitors". (In a later part of this enquiry it will be pointed out that there are no even approximately pure inductors.)

This statement needs some explanation. There is no circuit element that satisfies completely all (or perhaps any of) the tests that will be prescribed in their most stringent form. But, if elements are selected that very nearly satisfy all the tests, certain simple laws are found to be very nearly true of them; departures from the laws are concealed by experimental error in all but the most accurate experiments, and even in those are distinguishable from experimental error only by careful analysis. Accordingly we define a pure element as one of which these simple laws are true, and attribute departures from purity in actual elements to a combination of elements of different kinds, each of which would be pure in isolation. We find that the behaviour of actual elements can be explained on this hypothesis. The statements that follow must be interpreted in view of this procedure which, of course, is adopted with minor modifications in many branches of physics.

The distinction between pure resistors and pure capacitors is that the former do, and the latter do not, pass current in a D.C. circuit. (Pure capacitors thus violate the most fundamental property of D.C. circuits, namely that the path of the current lies wholly in conductors.) One test of purity applicable to both sub-classes is that the laws of addition in parallel and series must actually be obeyed. This test, since it involves a plurality of elements, can be applied only to the sub-class as a whole, not to individual members of it. It might fail completely for this reason; for impure elements might be additive, if they were all impure in the same way. Actually the test is very useful, because impure elements usually differ in their impurity. Further, the test, applied to capacitors, limits the sub-class to what we shall call (for a reason that will appear presently) "closed" capacitors (in common parlance, screened condensers). Other elements that are called pure capacitors according to convention—which we shall follow—are not additive.

Accordingly it is desirable to have other tests. Another class test is that the equivalence of two elements of the same sub-class should be independent of the indicator (i.e. of the nature of the mean it indicates) and of the generator (i.e. of its frequency and wave form). But this test, though satisfied by pure capacitors, is not fully satisfied by pure resistors; in ordinary language, their resistance varies slightly with the frequency.

There is a very perfect individual test for pure resistors. It is that the current through the resistor and the voltage across it, measured by cathode-ray oscillographs, must be "in phase"; in particular, the maxima and minima of current must be respectively simultaneous with the maxima and minima of voltage; this test can be applied without measuring anything, and involves only the relations of greater and less. The corresponding test for pure capacitors is less easily applied and its consideration therefore is postponed. The best individual test for capacitors is that which distinguishes them from resistors, namely that they pass no

direct current; but this is not conclusive, since it does not include dielectric loss and series resistance or inductance. The class tests for capacitors are therefore more important than the class tests for resistors.

The magnitudes characteristic of pure resistors that are additive in series and parallel turn out to be the same as D.C. resistance and conductance, at least at low frequencies; no separate name for them is required. On the other hand, new names are required for the corresponding magnitudes characteristic of capacitors; we shall call them elastance and capacitance. In modern terminology these magnitudes are distinguished from impedance and admittance; but, from our present standpoint, when we are considering independent measurement alone, they are the same magnitude according to our meaning of that expression (Part I, p. 645). For, whatever the frequency and wave-form, the ratio of the impedances or admittances of two pure capacitors is the same as the ratio of their elastances or capacitances.

The measurement of capacitance by the parallel connection of pure closed capacitors plays as important a part in the most refined calibration of pure capacitors as does series connection in the calibration of D.C. resistors. The measurement of capacitance is slightly less accurate than that of resistance, because (for a reason that will appear presently) the conditions in which the law of addition is true cannot be realized so precisely. For the same reason, and perhaps for others also, elastance plays a much less important part than capacitance in actual measurement; but it should be noted that, in so far as it is an independently measurable magnitude, it must be the reciprocal of capacitance, if the same standard is used to define both units.

§ 3. CAPACITANCE AS A DERIVED MAGNITUDE

If we measure by cathode-ray oscillographs both the current I flowing through a pure capacitor and the voltage V across it, then we find that I is proportional to the simultaneous dV/dt , and that this relation is independent of the particular generator used and of the way in which V varies. The constant of proportionality depends on the capacitor, and is therefore a derived magnitude; we may write the law

$$I = S \cdot C' \cdot dV/dt \quad \dots\dots (3.1)$$

C' turns out to be the same magnitude as the capacitance measured by addition, when the capacitor is closed and obeys the law of addition. If we choose suitably, with respect to the units of current, voltage and time, the closed capacitor to which unit capacitance is to be assigned, the scale factor S may be made 1. This is the convention always adopted; we shall assume its adoption in what follows.

Equation (3.1) is true even if the capacitor is not closed, so long as it satisfies the tests of purity other than that of addition. Accordingly, as indicated already, we shall regard capacitors obeying (3.1) as pure, even if they are not closed. It follows from (3.1) that, if there is a definite phase difference between I and V —which implies that both of them are sinusoidal—then it is 90° . This corresponds to the phase test for pure resistors; and since a phase difference of 90° implies that the maxima of one quantity coincides with the minima of the other, can be applied without measurement. But if the capacitor is not entirely pure, and the phase difference not exactly 90° , the distinction between a leading and a lagging current,

which will turn out to be important, will arise. It is therefore doubtful whether the test can actually be applied in a form in which it differs from the test whether (3.1) is true.

(3.1) cannot be established with great accuracy by oscillograph measurements, if only because the deduction of a derivative from a set of disparate values is always liable to considerable error. But on the assumption that (3.1) is true, it is possible to devise bridge networks fed by generators of sinusoidal wave-form (so that the relation between V and dV/dt is known) that permit C to be determined, both relative to other C 's and to the conductances of pure resistors. The consistency of the results obtained with such networks confirms with high accuracy the truth of (3.1), and the identity of C in that law with capacitance measured by addition in closed capacitors.

Just as the most refined measurement of resistance depends both on the law of addition and on Ohm's law, so the most refined measurement of capacitance depends both on addition and on (3.1). Bridges are also useful in judging equality in independent measurement; it is worth noting that, when they are used for this purpose, the wave-form of the generator need not be sinusoidal.

§ 4. CAPACITANCE AND GEOMETRICAL FORM: PERMITTIVITY

On examination of their structure, capacitors prove to consist essentially of a pair of conductors ("plates") separated by a gap filled with a non-conducting material, or dielectric and generally narrow compared with the dimensions of the plates. When the plates approximate sufficiently nearly to concentric spheres, coaxial infinite cylinders or infinite parallel planes, experiment shows that there are simple relations between capacitance and geometrical form. In the usual notation they are

$$C = K_1 \cdot r_1 r_2 / (r_2 - r_1) \quad (\text{spheres}) \quad \dots\dots (4.1)$$

$$C = K_2 / (\log r_2 - \log r_1) \quad (\text{cylinders}) \quad \dots\dots (4.2)$$

$$C = K_3 A / d \quad (\text{planes}) \quad \dots\dots (4.3)$$

K_1, K_2, K_3 vary with the dielectric; each is therefore a derived magnitude requiring a scale factor. But it turns out that they all vary together, so that—within experimental error— $K_1 = 2K_2 = 4\pi K_3$. Accordingly, if we write

$$K_1 = S \cdot \kappa; \quad K_2 = \frac{1}{2} S \cdot \kappa; \quad K_3 = 1/4\pi \cdot S \cdot \kappa. \quad \dots\dots (4.4)$$

a single magnitude κ (permittivity) characteristic of the dielectric and a single scale factor will suffice. Experiment shows further that the independence of permittivity and geometrical form is complete, so that if a value for κ is assigned by convention to one dielectric, it can be determined for any other by measuring the capacitance of a capacitor of any form, filled first uniformly with the standard dielectric and then uniformly with the other.

Somewhat more complicated formulae have also been established experimentally when the dielectric, though not uniform, is distributed in uniform layers. They involve different κ 's for the different layers, but they need no special consideration here.

All these formulae are special cases of a general rule for determining capacitance from the geometry of the capacitor and the nature of the dielectric. The form in

which it is given here depends historically on laws not mentioned so far. But the discovery of the rule by guesses based on empirical laws such as (4.1), (4.2), (4.3) would have been no more remarkable than the discovery of the Ampère-Neumann rule (Part I, (11), (11.1)) by a similar process.

The rule depends on a geometrical theorem that can be stated with sufficient accuracy and generality for our purpose thus* :—

Let, $s_1 \dots s_n$ be a set of closed non-intersecting surfaces. Let each point outside these surfaces be associated with a quantity κ , characteristic of it. Then it is possible to find a variable v having the following properties :—

- (1) v is constant over each surface s ;
- (2) at all points outside the surfaces s , v is continuous and finite or zero, being constant at all points sufficiently distant from the surfaces;
- (3) at each point at which κ is continuous, v satisfies the differential equation

$$\frac{d}{dx} \left(\kappa \frac{dv}{dx} \right) + \frac{d}{dy} \left(\kappa \frac{dv}{dy} \right) + \frac{d}{dz} \left(\kappa \frac{dv}{dz} \right) = 0; \quad \dots\dots (4.5)$$

- (4) at any surface at which κ changes discontinuously, $\kappa \cdot dv/d\mathbf{n}$ has the same value on either side of the surface, \mathbf{n} being the normal to it.

It follows from the properties of v that, if determinate values v_1, v_2 are assigned to two of the surfaces s_1, s_2 , while the values for the other surfaces are left undetermined, the quantity

$$\kappa \cdot dv \cdot d\mathbf{n} \cdot dS \quad 4\pi\phi_s, \quad \dots\dots (4.6)$$

where dS is an element of surface, \mathbf{n} is the outward normal, and the integral is taken over the surface, has the same modulus but opposite signs for the surfaces s_1, s_2 , and is zero for any other of the surfaces.

In applying this purely geometrical theorem to a capacitor, s_1, s_2 are identified with the surfaces of the plates, each plate including all the conductors connected conductively to it. v_1, v_2 are chosen arbitrarily. The value of κ associated with each point is the permittivity of the medium occupying it. If any point is occupied by a conductor insulated from both plates, κ for it is infinite; v is then constant over the surface of that conductor and ϕ_c for it is zero. Then v is in general determined everywhere; the quantity ϕ_s is everywhere proportional to $(v_1 - v_2)$; and the rule for capacitance is

$$\frac{\phi_c}{S} \quad \dots\dots (4.7)$$

where ϕ_c refers to the plates and S is the scale factor of (4.4).

This is the most general relation between capacitance and geometrical form; formulae (4.1), (4.2), (4.3) are special cases, but they do not provide a satisfactory test of it, because the capacitors to which they refer cannot be realized fully. However, they can be realized approximately; the rule is then found to give a "correction" that agrees with measurement. If the rule is applied to capacitors

* The reader will note that it is assumed that there are no surface or volume charges in the dielectric. The modification required if there are such charges is considered below in § 16. But it should be pointed out that the same assumption is made in § 5. He may note further that no reference is made to the vector D , of which ϕ_c is the flux. The reason for this omission will appear in later parts of this enquiry.

differing widely from these simple forms, mathematical difficulties often prevent the prediction of a definite value with which measurement can be compared. The direct evidence for the accurate and universal validity of the rule is therefore not complete; but it is very strong—stronger than that for any other “electrostatic” relation.

§ 5. THE FIELD EQUATIONS: ELECTRIC FIELD STRENGTH

In § 4 v is a mere mathematical variable, convenient (but perhaps not absolutely necessary) in arriving at formulae (such as (4.1-3)) applicable to experiment; it does not appear at all in the formulae. But further inquiry shows that v , or more accurately $\text{grad } v$, has great physical significance.

In Part I the magnitude H was introduced to express the fact that there is a group of phenomena, including the deflection of a compass needle or of a cathode ray, that are characteristic of the space about a current-carrying circuit, and are determined by the product of the current in that circuit by a geometrical function G . There is another group of phenomena similarly characteristic of the space about conductors between which voltages are maintained; like the “magnetic phenomena”, they form a group, because if one of them is the same at two points in such a space (or spaces), then each of them is the same at those points. The group includes again the deflection of a cathode-ray; it includes also the Kerr electro-optical effect and the breakdown of a dielectric.

Suppose that we identify the surfaces of the conductors with the surfaces s , assign values of v to them so that $v_k - v_l$ is the voltage V_{kl} maintained between s_k and s_l , and calculate v throughout the space between them by means of the theorem of § 4. Then we find that the phenomena are determined by $\text{grad } v$ in the sense that they are the same wherever $\text{grad } v$ is the same. Accordingly

$$\mathcal{E} = \text{grad } v \quad \dots\dots(5.1)$$

is a defined magnitude determining these electric phenomena in the same way as H determines the magnetic phenomena, it may be fitly called the electric field strength. If there are only two conductors between which a voltage is maintained, if V is that voltage, and if the dielectric is uniform, the similarity between E and H may be made even more evident by writing

$$\mathcal{E} = VF, \quad \dots\dots(5.2)$$

where F , like G , is a function depending only on the geometry of the system.

In view of these facts we shall term the propositions of § 4 collectively the *field equations*. Further, we shall assume—for this is obviously the easiest way—that the v 's are to be assigned to the conductors in accordance with the voltages, by assigning $v=0$ to one conductor (zero plate) and to the others their voltages to this conductor.

In some circumstances—it is unnecessary here to inquire what they are—it is possible to establish that the voltage between the zero plate and a probe introduced into the space between the conductors is equal to v , calculated from the field equations. Then v , as well as $\text{grad } v$, has physical significance. Measurements of probe voltages are inaccurate and have little evidential value. On the other hand, \mathcal{E} can be determined with considerable accuracy from the deflection of a cathode-ray, by the use of a theorem that will be examined in § 15. It is possible that the

future development of electron optics may provide evidence for the field equations more stringent than that derived from the measurement of capacitance; but for the present they must be based primarily on such measurement.

v of this section is closely related to the v that was introduced in Part I to explain the properties of conducting media. It follows from the field equations and from the properties of v set forth in § 13 of Part I that, if the dielectric of a capacitor is replaced by a medium whose conductivity is everywhere proportional to the permittivity of the dielectric, and if the plates are still maintained at the same voltage, then the v of the conductor is everywhere the same as the v of the capacitor. This relation is sometimes used to determine the distribution of v in a capacitor of complicated form. But it is more important because there is no sharp distinction between dielectrics and conductors, there are media in which it would be possible to determine both equipotentials as indicated in § 13 of Part I, and the electric field strength by one of the group of phenomena mentioned earlier in this section; more generally we find that the properties of most dielectrics can be explained by attributing to them both permittivity and conductivity. It is an experimental fact of great practical importance that, in isotropic media, the distribution of v is the same whether it is derived from conductive or from electrostatic phenomena. Perhaps the strongest evidence on this point is the well-known fact, frequently established with great precision, that the "loss tangent", and therefore σ/κ , for any uniform medium is independent of geometrical form.

§ 6. CLOSED CAPACITORS: SCREENS

The exceptions to the statement in § 4, that v is determined everywhere by the assignment of v_1, v_2 , are very important.

(a) If a conductor, insulated from both plates, encloses one but not the other, then v is completely indeterminate unless something is added to the rule. The obvious addition, based on the fact that ϕ_e for a conductor insulated from both plates is zero, is that, if ϕ'_e is the value of ϕ_e for the enclosed plate, then $-\phi'_e$ is its value for the inner surface of the enclosing conductor, $+\phi'_e$ for its outer surface, and $-\phi'_e$ for the unenclosed plate. The experimental facts that justify this addition will be stated later. With this addition, which will be assumed in what follows, v in this case is determinate everywhere.

(b) If one of the plates completely encloses the other, then the assignment of v to the enclosed plate and to the inner surface of the enclosing plate, though it determines v everywhere within the enclosing plate, does not determine v anywhere outside it. The corresponding experimental fact is that changes in the voltages between conductors inside the enclosing plate (including the inner surface of that plate) produce no change of voltage between any pair of conductors outside the enclosing plate (including the outer surface of that plate), and no change of the field strength outside that plate; and conversely, interchanging "inside" and "inner" with "outside" and "outer".

This deduction from the field equations can be tested with great sensitivity; since it is true, some variations from the rule of capacitance given by the field equations as stated in § 4, which might otherwise be possible, are excluded. But the fact is even more important practically than theoretically. Thus, it is used in radio-frequency experiments when the observer, in order to isolate himself from

irrelevant disturbances, encloses himself and all his apparatus inside a metallic cage. (In view of later discussions, it is well to point out here that the efficiency of the cage for this purpose does *not* depend on its being connected to any other conductor, whether inside or outside the cage.)

Again, it makes it possible to limit the conductors that have to be taken into account in any electrostatic problem to a definite group. If one of a group of conductors encloses all the rest, then any conductor outside the group—and outside the enclosing conductor—can be ignored. This limitation and isolation by a “screen” is necessary both in theoretical arguments and in all experiments except the very roughest. Accordingly it will always be assumed hereafter, unless the contrary is stated explicitly, that the system under discussion is a group within and including a surrounding screen. Of course an isolating screen need not in practice be absolutely complete; the metallic cage above mentioned is often made of wire gauze or netting. The effect of a small aperture tends to zero with the ratio of its area to the whole area of the screen, and is often quite inappreciable.

Other closely related deductions from (b) are that the capacitance of a capacitor of which one plate (screen) completely surrounds the other is independent of any body outside it, and that the capacitance of a pair of such capacitors, with their screens and their plates respectively connected together, is the sum of their individual capacitances. If one plate does not surround the other, capacitance will not in general be additive in parallel, because the field between the plates of one capacitor affects the field between the plates of the other. This is, of course, why capacitors that obey the laws of addition were called “closed” in §2.

If a capacitor could not be “closed”, and obey the laws of addition, unless the screen *completely* surrounded the plate, a closed capacitor would be an unattainable ideal. For, in order that the plate of a screen-and-plate capacitor should be connected to anything outside the screen, e.g. another plate, there must be an aperture in its screen through which a lead may pass. The aperture itself, if small enough, may be unimportant; but the lead passing through it is another matter. It is a question of fact, not to be decided by anything stated so far, whether the departure from ideal closure, represented by the passage of a lead through the screen, produces a failure of the law of addition.

The answer depends, of course, on the sensitivity of the criterion of equality and on the size and location of the lead. If the most sensitive means for judging equality is used, and the capacitors have ordinary terminals, then the law of addition is found to be not *generally* true, although it is true for capacitors exceeding (say) $0.5\ \mu\text{F.}$ in value, provided that reasonable precautions are taken in choosing and arranging the leads. It is, however, possible so to design capacitors that the leads connecting them together make no detectable contribution to their capacitance. For example, the plate terminals may take the form of sockets mounted in small apertures in the screen, so that the face of the socket is flush with the surface of the screen. If the sockets of two such capacitors register with one another when the screens are placed in contact, then they can be connected together by a double-ended plug, which is inserted in one socket before the capacitors are brought together, and which becomes totally enclosed when the connection is complete. When capacitors of this type are used, the law of addition can be verified with the highest accuracy over a very large range of values; indeed, it is probably the most

accurately known of all the quantitative laws of electrostatics. Moreover (see § 12 below) there are methods of assigning capacitances to ordinary standard capacitors that are also additive in respect of the most sensitive criteria of equality; and these capacitances agree within experimental error with those calculated from the rule of capacitance. Accordingly there are actual capacitors completely equivalent to ideal closed capacitors, having no leads through their screens; and in basing the laws of electrostatics on the properties of ideal closed capacitances, we shall not be departing from our principle of relying only on real experiments and real conceptions.

There is a corresponding deduction from (a), namely that elastance is additive in series if, and in general only if, the plate common to the two capacitors completely encloses the other plate of one of the capacitors, and is completely enclosed by the other plate of the other. This condition is not easily realized, even if the leads are ignored, and elastance, even if it proved accurately additive in the presence of leads, would not be important. Actually the departures from accurate additivity of elastance are of the same nature as those just discussed; they are often concealed by experimental error; and even when they are not concealed, they may be abolished by the methods of § 12 below.

§ 7. ELECTROMETERS

We now return to the main theme, namely the coordination of the laws of current circuits.

We note that electrostatic voltmeters and electrometers are essentially closed capacitors in which the plate is movable relative to the screen, so that their capacitance varies with the motion of the plate, and the voltage V to be measured is applied between plate and screen. (If such voltmeters were used to investigate (3.1), account would have to be taken of the variation of their capacitance, which is part of C , with V .) Experiment shows that their behaviour is consistent with the law

$$F_q = S \cdot dC/dq \cdot V^2, \quad \dots\dots(7.1)$$

where F_q is the force tending to increase the coordinate q , the voltage being maintained independent of q by means of a battery or generator, and S (which is positive) is the same for all voltmeters that obey (7.1), so that it is a scale factor that does not need to be supplemented by a derived magnitude.

The only voltmeters in which the law can be established accurately are those in which C can be calculated accurately from the field equations; but, in some of these, it can be established as accurately as any of the laws previously mentioned in this Part. By assigning values to S in (4.4), (4.7), (7.1), and to κ of a standard dielectric, the units of C , V , I , κ can be fixed in terms of the units of length, area, time and force. They could actually be fixed by this method with an accuracy little less than that with which they are now fixed by other methods. In "electrostatic absolute units", κ is made unity for vacuum, S of (4.4), (4.7) are made unity, and S of (7.1) is made $\frac{1}{2}$. (The reason for this last choice will appear presently.)

§ 8. THE CHARGE IN A CAPACITOR

We have said that (7.1) is true only if V is made independent of q by connecting plate and screens through a battery or generator. Electrometers are sometimes

used idiostatically with plate and screen insulated from each other; then V varies with their relative position. We shall now show that the law corresponding to (7.1) for this case can be inferred from the laws already stated.

Suppose that a direction-sensitive ammeter is permanently connected in series with one of the terminals of the capacitor. While the capacitor is a circuit element, let us record I , the current flowing through the ammeter, as a function of time; and let us define a quantity Q , the charge in the capacitor, by

$$Q = \int_0^V I \cdot dt, \quad \dots\dots(8.1)$$

where the limits of the integral indicate that it is to be taken between an instant when the voltage across the capacitor is zero and an instant when it is V . Then from (3.1), with $S = 1$

$$Q = CV. \quad \dots\dots(8.2)$$

Comparison of (4.7) and (8.2) shows that

$$|Q| = |\phi_e|, \quad \dots\dots(8.3)$$

accordingly in so far as Q , defined by (8.2), receives physical significance from the considerations of this section, ϕ_e receives the same significance.

When the terminals of a pure capacitor are not connected in a circuit, $I = 0$. If (3.1) is true of these circumstances (an assumption not used so far), then V , and therefore by definition Q , must remain constant. Experiment shows that this condition may be approached very nearly in suitable circumstances; any change in V can be attributed to the presence of a conductor between the terminals whose conductance (representing a slight impurity) is so small that it cannot be detected in the range of frequencies within which C is most conveniently measured.

Let us now connect the terminals of the charged capacitor by a resistor of resistance R . Then, if Ohm's law is true when the appearance of a voltage across the resistor is not associated with the presence of a generator in the circuit of which it forms part (again a new assumption), we must have

$$I = -V/R, \quad dI/dt = (-1/R) \frac{dV}{dt} = -I/CR \quad \dots\dots(8.4)$$

$$\int_0^V Idt = \int_{I=0}^{I=-V/R} -CRdI = CV. \quad \dots\dots(8.5)$$

This relation is confirmed by experiment, i.e. Q is the same whether it is estimated by charging or by discharging, and our assumptions are therefore justified. From these experiments we conclude that Q or ϕ_e is a significant magnitude, characteristic of the state of a capacitor, having a constant terminal-voltage, when its terminals are insulated.

We expect then that, when the terminals of the capacitor are insulated, and its configuration changes, Q will still remain constant. If so, in order to ascertain the behaviour of the electrometer when its terminals are insulated, we may substitute for V in (7.1) in terms of Q from (8.2), and perform the differentiation with respect to q subject to the constancy of Q . We thus have

$$F_q = -S \cdot \frac{d}{dq} \left(\frac{1}{C} \right) \cdot Q^2. \quad \dots\dots(8.6)$$

This law is confirmed by measurements; but, owing to the difficulty of maintaining the terminals truly insulated, it cannot be established as accurately as (7.1). Its truth is, however, a great part of the evidence for the propositions about Q .

§ 9. ENERGY IN A CAPACITOR

(8.6) is the law that would follow from general dynamical principles, if it were known that a pure insulated capacitor is a conservative system and that its potential energy is $\frac{1}{2}SQ^2/C$. We therefore inquire whether there is any other support for these suggestions, taking into account equation (6) of Part I.

From (3.1)

$$\int_{t_1}^{t_2} IV dt = \frac{1}{2}C(V_2^2 - V_1^2), \quad \dots\dots(9.1)$$

where V_1, V_2 are the voltages at t_1, t_2 . It follows that, even if (6) were true of the instantaneous power converted into heat in a pure capacitor, the total energy and mean power dissipated in a complete cycle or in any number of cycles, would be zero. Experiment shows that it is zero or, more accurately, that any power dissipated in a capacitor in such circumstances can be associated with a departure of the capacitor from complete purity that can be detected in other ways. But it shows also that, in a pure capacitor, no energy is dissipated in even part of a cycle throughout which IV is of the same sign, although then, if (6) were true, power would be drawn from the generator and dissipated. Moreover, when a charged capacitor is disconnected from the generator by which it has been charged and is discharged through a resistor, I' and I being observed during the discharge, power is dissipated in the resistor according to (6), being apparently drawn from the capacitor.

All this is consistent with the storage in the charged capacitor of potential energy W_Q , where, if the units of I, V are chosen to make the scale factor in (6) unity,

$$W_Q = \frac{1}{2}Q^2/C. \quad \dots\dots(9.2)$$

Accordingly (8.6) will be identical with the law to be expected on general principles, so long as the units are chosen so that $S = \frac{1}{2}$. This is the reason for that choice in (7.1).

The potential energy can also be written

$$W_1 = \frac{1}{2}C'V^2. \quad \dots\dots(9.3)$$

But it must *not* be concluded that the force tending to increase q , when V is constant, is $-dW_1/dq$. For, when V is constant, the capacitor alone is not a conservative system, because it can exchange energy with the generator by which V is maintained. The fact that F_q in these circumstances is given by (7.1) shows that the energy drawn from the generator in maintaining V is twice as great as, and opposite in sign to, the change in potential energy.

§ 10. GENERALIZATION OF CAPACITANCE AND CHARGE

The capacitors that we have considered so far ("simple" capacitors) are essentially pairs of conductors. Any conductor other than the pair that is relevant at all, in the sense that its removal would effect the capacitance of the pair, must be insulated from the pair, so that it is effectively part of the dielectric. But

sometimes we have to consider systems of more than two conductors, each pair of which is, or may be, maintained at a different non-zero voltage. The main, if not the only, examples of such systems offered in the older text-books are electrometers used heterostatically; but today more important and interesting examples are furnished by telephone cables and thermionic-valve circuits.

A study of the field equations suggests that the magnitudes *capacitance* and *charge* may have some relevance to such systems. Thus, suppose that we have N conductors, distinguished by suffixes $1, \dots, m, \dots, n, \dots, N$, of which one is a screen surrounding all the rest; and that we consider the simple capacitor of which one plate is the conductor 1, characterized by v_1 , and the other all the remaining conductors, characterized by v_0 . If ϕ_m is the integral of (4.6) taken over m , and C_1 is the capacitance of this simple capacitor, then we must have

$$\phi_1 = -\sum \phi_m \quad \dots\dots(10.1)$$

$$C_1 = \frac{\phi_1}{v_1 - v_0} \quad \phi_m \quad \dots\dots(10.2)$$

where the summation excludes 1; and, if we define C_{1m} by

$$C_{1m} = \left| \frac{\phi_m}{v_1 - v_0} \right| \quad \dots\dots(10.3)$$

we must have

$$C_1 = \sum C_{1m} \quad \dots\dots(10.4)$$

Suppose now that we divide the conductors differently, taking 2 as one plate and all the others as the other plate. The capacitance C_2 of this simple capacitor will have components C_{2m} defined analogously to (10.3). By selecting in turn each of the conductors as one plate of the simple capacitor, we can find all the C_{mn} . It is now found to be a consequence of the field equations that

$$C_{mn} = C_{nm} \quad \dots\dots(10.5)$$

A more general consequence is that, if we assign to the conductors the values $v_1, \dots, v_m, \dots, v_n$, which may all be different, then, if ϕ_m and C_{mn} are defined as before,

$$\phi_m = \sum C_{mn}(v_m - v_n), \quad \dots\dots(10.6)$$

$$\sum \phi_m = 0, \quad \dots\dots(10.7)$$

where the summations extend over all the conductors. If we write $Q_n = \phi_n$, $v_m - v_n = V_{mn}$, these become

$$Q_m = \sum C_{mn} V_{mn}; \quad \sum Q_m = 0. \quad \dots\dots(10.8)$$

This clearly suggests that the C_{mn} (which will be termed *mutual capacitances*) have the same physical significance as capacitances, and that, if the V_{mn} are identified with the voltages between mn , the Q_m have the physical significance of charges. Indeed, the system of conductors should be equivalent to a set of terminals $1, \dots, m, n, \dots$, interconnected by simple closed capacitors of capacitance C_{mn} . The only difference between the system and a set of terminals interconnected by real simple closed capacitors is that each C_{mn} is determined, not by the configuration of two conductors only (namely the plate and its screen), but by the configuration of all the conductors.

Further, we should expect each Q_m to be equal to some integral $\int I dt$, where I is a current that flows to or from it during its charging or discharging. It is unnecessary for our purpose to define these relations generally; but it may be noted that Q_m should be $\int_0^\infty I dt$, where I is the current that flows in a circuit connecting the charged m to all the other conductors connected together. However, one matter requires more detailed attention. Hitherto, when Q has been the charge in a capacitor, it has been unnecessary to pay attention to the sign of Q , and the sign of I in the integral to which it is equal. Now Q_m is the charge on a single body, and, as (10.8) shows, it must sometimes be of one sign and sometimes of the other; it is important to lay down rules by which its sign can be ascertained. Formally, in order to solve all problems, these rules should be given in terms of the red and black terminals of ammeters and voltmeters discussed in Part I. But, since there is really no difficulty in the matter for anyone practised in electrical experiments, it suffices to say that the charge in a simple closed capacitor is to be regarded as positive if the voltage of the plate relative to the screen is positive, and that, when we write

$$I dt, \quad \dots\dots(10.9)$$

I must be the positive current flowing during discharge from the plate to the

§ 11. PROPERTIES OF A COMPLEX CAPACITOR

All these propositions about a complex capacitor (for so we shall term a set of conductors such as is considered in the preceding section) are so far mere expectations, based on suggestions derived from the form of the field equations. There is a major difficulty in relating any of them to experiment. It is that, while the reasoning that leads to the propositions involves the assumption that the conductors are all isolated, any possible experiment requires leads between them. However, if we suppose that the presence of the leads does not destroy the equivalence between a complex capacitor and a set of terminals connected by simple closed capacitors, but merely modifies the mutual capacitances represented by the capacitances of those simple capacitors, then some experimental tests can be applied.

Thus, the propositions indicate that the formula corresponding to (9.3) should be

$$W_r = \frac{1}{2} \sum C_{mn} V_{mn}^2, \quad \dots\dots(11.1)$$

where the summation is taken over all pairs without permutations. Since

$$V_{mn} = -V_{nm}; \quad V_{mo} = V_{mn} + V_{no} \quad \dots\dots(11.2)$$

there are only $N-1$ independent V_{mn} ; by solving the $N-1$ independent equations (10.8), we can obtain them as linear functions of the Q_m . Substituting these values in (11.1), we have

$$W_Q = \frac{1}{2} \sum p_{mn} Q_m Q_n, \quad \dots\dots(11.3)$$

where the p_{mn} are functions of the C_{mn} subject to certain interrelations that need not be considered here. In an electrometer used heterostatically with insulated

terminals, F_q should be $-dW_q/dq$; accordingly (11.3) predicts certain laws relating the forces on the electrometer to the charges on its parts, in which the constants are functions of the mutual capacitances. It is possible, as in many similar instances in various branches of physics, to apply tests that assume merely that these constants are indeed constants, and that do not require any knowledge of their actual values; if these tests are satisfied, some evidence will be obtained for the conclusions of this section without any detailed knowledge of the mutual capacitances.

Qualitative tests of this nature applied to electrometers are partially, but not completely, satisfied; it is almost certain that the discrepancies arise, not from any failure of the conclusions of this paragraph, but from the difficulty of taking into account *all* the mutual capacitances and of otherwise fulfilling exactly the conditions postulated. Perhaps better evidence of this nature is derived from valve-circuit theory, where again important conclusions can be reached, independent of accurate knowledge of the mutual capacitances. But the accuracy possible in all such tests is very low.

§ 12. MEASUREMENT OF MUTUAL CAPACITANCE: SCREENING

However, if a set of conductors is indeed equivalent to a network of closed capacitors, representing the mutual capacitances as modified by the leads, the capacitances of these capacitors can be measured by apparatus of the kind shown in figure 1.

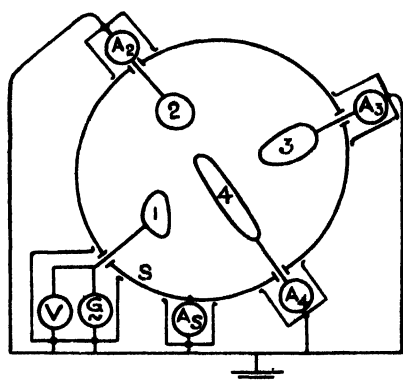


Figure 1 Scheme for the measurement of the mutual capacitance C_{mn} of any pair of conductors m, n .

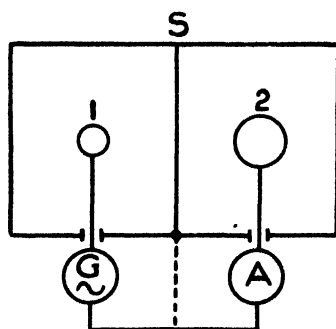


Figure 2.

Here 1, 2, 3, 4 are conductors within the screen S . G is a generator with one terminal connected to 1; A_2, A_3, A_4, A_S are ammeters with one terminal connected respectively to 2, 3, 4, S ; all the connecting leads are as thin as possible. The other terminals of the generator and ammeters are all connected together. The ammeters are all of low impedance, so that the voltage between 1 and all the other bodies is V , measurable by a voltmeter across the generator, and there is no voltage between any pair of these bodies.

If the generator and instruments had no effect beyond supplying and measuring current and voltage, it would be desirable to place them within S , in order that the

leads should be as short as possible. Examination of the equivalent network would then show that the only current I_m through A_m is the current through the capacitor C_{1m} ; consequently we should have

$$I_m = C_{1m} \cdot dV/dt. \quad \dots\dots(12.1)$$

Since I_m is measurable by A_m and V by the voltmeter, C_{1m} could be determined. (10.4) would be necessarily true in virtue of Kirchhoff's First Law and the method of measurement. C_{2m} etc. could be measured by substituting the other bodies in turn for 1; but experiment alone could determine whether $C_{mn} = C_{nm}$.

But the hypothesis is not true. The generator and instruments are really part of the system, having mutual capacitances with each other and with 1, 2, 3, 4, S. In order that (12.1) should be true in spite of these mutual capacitances, the generator and ammeters are placed outside S, and connected to the conductors within S by leads passing through apertures in S; and each is surrounded by its own screen connected to a common terminal (see § 14 below). This is a simple example of the art of screening, which forms an important part of A.C. technique. We shall not discuss it fully here; for one of us has discussed it in a recent publication (Hartshorn 1940). But some statement of the underlying principle is desirable.

The principle involves more than that of the metallic cage mentioned in § 6. The object of the screens is not to isolate the generator and instruments completely from each other and from bodies inside S; but, firstly, to limit the extension in space of the electric fields of the bodies within the several screens and, secondly, to direct the capacitance currents associated with those fields to appropriate terminals of the equivalent network; or, in other words, not so much to abolish capacitance currents as to direct unwanted capacitance currents into paths where they are innocuous. This function can be explained with reference to figure 2.

The introduction of any conductor S between two others 1, 2, will decrease the mutual capacitance C_{12} to (say) C'_{12} ; but it will also introduce two new mutual capacitances C'_{1s} , C'_{2s} . The total capacitance of the network between 1, 2 will become $C'_{12} + C'_{1s}C'_{2s} / (C'_{1s} + C'_{2s})$, and this will always be greater than C_{12} . If, as in figure 2, S surrounds both 1 and 2, C'_{12} will be very small and tend to zero with the area of the aperture in S; but the total capacitance will still be greater than C_{12} in the absence of S. Accordingly S in itself increases rather than decreases the mutual reaction of 1, 2. But if, as shown by the dotted lines, S is connected to the common terminal of the generator and ammeters, the current through C'_{1s} is diverted from the ammeter to the dotted connection, and the current through C'_{2s} is abolished, because there is no voltage driving it.

The action of the screens in figure 1 can be understood in the light of this simple example. But that example is merely an illustration of the principle involved; it must not be supposed that any feature of figure 1 is characteristic of all applications of the principle. Thus, it is not always desirable that each element should have its own screen, or that all screens should be connected to a common terminal. The application of the principle is so complex that it is scarcely possible to state any practical rule that is valid without limitation.

Experiments are not often made in exactly the way indicated in figure 1; but the various bridge and resonance methods for the measurement of mutual capacitances often (called "direct" or "partial" capacitances by communication engineers, who are chiefly concerned with them) are extensions of the same

principle, and depend upon (12.1) just as bridge methods for simple capacitors depend upon (3.1). These methods lead to consistent values in accordance with §§ 10, 11; in particular, such experiments fully establish the fact that C_{mn} , though dependent on the configuration of all the conductors, is not dependent on their voltages. The relation $C_{mn} = C_{nm}$ can also be established with high accuracy, provided that, in performing the experiment, the generator connected to m and the ammeter connected to n have their positions in the network interchanged without any disturbance of the leads that penetrate S . The values obtained often agree well with those calculated from the field equations. The measurements made by Rosa and Dorsey (1907) on the complex capacitor (a central cylinder with guarding extensions), used in their determination of ϵ , is an outstanding example. Such experiments provide the best evidence for the general theory of complex capacitors, i.e. in traditional language, "systems of conductors".

Finally, precision methods of measuring the capacitance of standard capacitors are essentially methods of measuring mutual capacitance. It is therefore impossible in principle that they should lead to entirely definite values. For, as we have seen, the mutual capacitance of two conductors depends on all the other conductors with which it forms a definite group in the sense of § 6; moreover, measurable mutual capacitances are always modified by the presence of leads; the very purpose of a standard capacitor forbids that it should always be associated with the same conductors or joined to the same leads. However, the margin of uncertainty can be made very small, e.g. 2 parts in a hundred thousand for a capacitance exceeding $0.1 \mu F$.; and, as indicated in § 6, the rules of addition in parallel are satisfied by some values within this margin.

In some exceptional circumstances it is possible to assign values to elastances within a comparable margin of uncertainty, and to show that elastance is additive in series. But it will not usually be possible to prove this with high accuracy by the use of standard capacitors; for, as usually constructed, they do not satisfy sufficiently nearly the condition, stated in § 6, for the ideal addition of elastance.

§ 13 ABSOLUTE CHARGE

In the foregoing discussion the charge on a body has appeared as a property that it possesses in virtue of its relation to other bodies with which it forms a capacitor. On the other hand, in the classical theory of electrostatics, charge appears as an "absolute" property of a body, independent of its relation to any other body, and moreover, as one that can be measured without the knowledge of any of the algebraic laws involving charge, and, in particular, of those algebraic laws that concern capacitors. We have to inquire on what facts such a conception of absolute charge is, or could be, based.

Suppose that a closed capacitor is charged by connecting the plate to the screen through a battery, and that the plate is then discharged by connecting it to the screen through an ammeter. If I is the current through the ammeter during the discharge, the charge acquired by the plate during the charging is given by

$Q = \int_0 I dt$. Suppose now that the plate, after having been charged, is removed from the screen in which it was charged and, remaining always insulated, is transferred to the interior of another screen. If it is now connected to this screen

through an ammeter, $\int_0^\infty I dt$ should have the same value as before, whatever the form of the second screen, wherever the plate is placed within this screen, whether the first screen is insulated from the second, or whether a voltage, zero or non-zero, is maintained between them by a conductor or battery. In classical electrostatics this proposition is stated as a consequence of the theorem that the charge on a conductor resides wholly on the outside of it; but it does not follow from anything in our previous discussion. If it is true, it must be established by direct experiment. If it proves to be the fact that, whenever a body, charged as part of a capacitor, simple or complex, is discharged to an enclosing shield, $\int_0^\infty I dt$ turns out to be the same as the charge that it acquired as part of the capacitor, then *charge* will have a significance independent of any particular capacitor, though still not independent of the general properties of capacitors.

The experiment would be difficult to perform with any accuracy, and the test has probably never been applied. But an equivalent test may be devised by noting that the law $Q = CV$ is inseparable, according to our exposition, from $Q = \int_0^\infty I dt$. If, instead of discharging the body to the second screen, we measure the capacitance of the closed capacitor, formed by the body and the second screen, and also the voltage across the capacitor, then we should find that, though C and V depend on the form of the second screen and on the position of the charged body within it, CV is independent of both and equal to the charge that the body acquires in the first screen.

However, the test is still not convenient, since it requires the measurement of a different capacitance each time a charge is to be measured. The celebrated "Faraday cage" is essentially a device for overcoming this difficulty. It is illustrated in figure 3, which shows diagrammatically the complete electrical system involved in any attempt to perform the experiment accurately. Here the internal screen 1 is the "cage" into which the charged body 3 is introduced, so that 1 and 3 form a closed capacitor of which the product VC is required. An outer screen 2 surrounds 1, and a voltmeter V_{12} is connected between them; this arrangement provides another closed capacitor whose capacitance C_{12} and voltage V_{12} are accurately measurable. Now the complete capacitor 3, 1, 2 has that special form for which elastance is strictly additive (§ 6), and which is therefore completely represented by two simple capacitors in series, C_{13} and C_{12} in figure 3. Thus, when this complex capacitor is charged, which is the case when the body 3 enters the cage 1, each of these simple capacitors receives the same charge. Consequently, if V_{12} is originally zero, the value that it will acquire when the body 3 bearing charge Q enters the cage 1 is given by

$$Q = C_{12}V_{12}, \quad \dots\dots(13.1)$$

where C_{12} includes the capacitance of the voltmeter. Accordingly a knowledge of C_{12} alone is required, and this is independent of the form of 3 and of its position within the cage.

It is now quite practicable to apply the test. 3 is given a known charge Q by charging a closed capacitor, of which 3 forms the plate and 4 the screen, to a

voltage V_{34} by a battery between 3 and 4. 3 is now introduced into 1. If V_{12} were originally zero, it should now be given by

$$C_{12}V_{12} = C_{34}V_{34}. \quad \dots\dots(13.2)$$

If this relation is always fulfilled, we have a way of measuring the charge on a body independently of the capacitor in which it received its charge.

V_{12} need not be zero originally. If it were not, then (13.1), (13.2) should be true, if δV_{12} is substituted for V_{12} , where δV_{12} is the change of V_{12} when 3 is introduced. This observation is important, because it leads to a method of measuring Q that does not require the prior establishment of methods of measuring C and V , or indeed of any other magnitude. Suppose that the voltmeters V_{12} , V_{34} are originally uncalibrated. We make some mark on the scale of V_{34} , and assert that, when its pointer stands at this mark, we are going to say that 3, being within 4, has the charge 1. We transfer 3 to within the cage, and, V_{12} having been originally zero, note the reading of V_{12} , and call this 1. We discharge 3 to the cage (by virtue

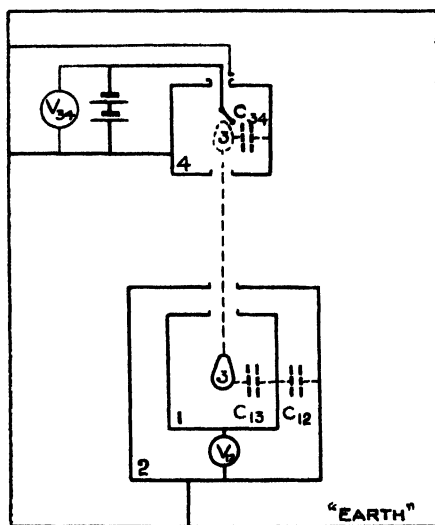


Figure 3. The "Faraday Cage" Experiment.

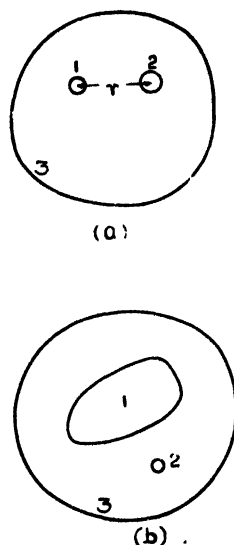


Figure 4. The problem of three charged bodies. Coulomb's Law is an approximation for case (a) and $F_r = -Q_2 \frac{dr}{dr}$ an approximation for case (b).

of principles already laid down, this should not change the reading of V_{12} and experiment confirms this), put 3 back into 4, charge it so that again it has charge 1, and transfer it again to the cage. The reading of V_{12} now we mark 2. Repeating the process, we get points on the scale of V_{12} marked 3, 4, In order to ascertain whether these marks on V_{12} provide a consistent scale of charge, we change the voltage V_{34} or the body 3, or both, until we find that, when 3 charged in 4 is transferred to 1, it causes V_{12} to change from 0 to 2. Then, repeating this process, we find that the second transfer changes V_{12} from 2 to 4. If the rules of addition, tested by a sufficient variety of such experiments, are always fulfilled, the calibrated

scale of V_{12} will measure charge, and we shall have established charge as an independently measurable magnitude; we shall have gone a long way to providing evidence for the fundamental postulate of classical electrostatics. Of course nowadays this independent measurement of Q is of little importance; it is easier and more accurate to measure Q as the derived magnitude CV . But it is remarkable that no experiment of this kind, so necessary as a starting point for classical theory, appears in either the standard treatises or the histories. However, this becomes less surprising when we observe that all the laws of physical addition, even those of length and mass, have almost always been accepted purely intuitively. They are so fundamental that the underlying assumptions are made unconsciously. It is, however, desirable that they should be brought to light.

But, though we have now dissociated Q from any magnitude characteristic of capacitors, we have not rendered it entirely independent of them. It still is significant only in connection with the charging of capacitors. Two further experiments are needed to establish complete independence. One arises from the discovery that bodies that have not been charged as part of a capacitor may produce a reading of V_{12} when they are introduced into the cage. In order to establish that the magnitude charge may fitly characterize such bodies, we must show that the introduction into the cage of n bodies, each of which when introduced singly gives the same value of δV_{12} , gives a reading of $n\delta V_{12}$.

For the second experiment, we require two similar bodies 3 as shown in figure 3, and we break the connections of 2 and 4 to earth, and connect these screens together. We then have in figure 4 a complex capacitor consisting of three simple closed capacitors in series, namely those shown dotted in figure 3. We connect the two bodies 3 by a battery in series with an ammeter, and measure $\int I dt$, where I is the current that flows during the charging of the complex capacitance. We now remove 3 from within 1, and introduce into the cage 1 the body 3 which has been charged within 4. We shall find that the charge on 3 so measured is $\int I dt$ measured during the charging. We have then proved that the transport of a charged body between the two positions 3 in figure 2 has precisely the same effect, in respect of this apparatus, as the passage of a current I between those two positions, so long as the charge on the body is $\int I dt$, and *vice versa*. This leads us to inquire whether the transport of a charge is equivalent to the passage of a current in respect of other apparatus, e.g. that for measuring the magnetic field produced by a current. We find that it is, so long as the path traversed by the charge during its transport is the same as the path of the current. When we have proved that, we have proved that charge has a significance wider than that of electrostatic experiments, all of which involve capacitors; we shall have proved, in the language of our ancestors, the "identity of all forms of electricity".

§ 14. "EARTH"

Earth is a conception hardly less fundamental in classical theory than *charge*; but the explanations offered of what constitutes earth* are usually very vague, and

* We leave on one side problems concerning the "atmospheric gradient" or old-fashioned telegraphy with an "earth return". For in them "earth" undoubtedly means The Earth. But it is perhaps worth noting that there was a period (roughly that of the 1870's) when electrical nomenclature was becoming standardized and telegraphy was the only form of electrical engineering. The present uses of "earth" are probably derived in some measure from "earth returns".

sometimes demonstrably false. In the absence of a formal definition the meaning of earth can be ascertained only by examining the statements that are made about it.

Thus *earth* is the thing to which zero potential is assigned in conventional discussions of "systems of conductors". Since differences of potential are voltages, it is the thing from which voltages are measured, i.e. the "zero plate" of § 5. Now if there is a terminal to which many more conductors are connected than to any other, it is clearly convenient to take that conductor as the zero plate. In apparatus containing many screens through which leads pass, and which are therefore connected together, such as that of figure 1, the terminal to which the screens are connected is such a terminal. If therefore *earth* means the common terminal of screens in such apparatus, it has a clear significance in the light of our exposition, which is all based on the conception of capacitance. Moreover, such a use of the term would agree with ordinary laboratory usage; for such a common terminal is called "earth" in that usage; indeed, the avoidance of *earth* in our discussion of screening must have appeared artificial to many readers. Further, the usage has strong justification; for in such apparatus the sheathing of cables almost always forms part of the screens, and this sheathing is always connected to The Earth.

If, however, the observer and all his apparatus are enclosed in a complete screen through which no leads pass, this screen will always be used as the common terminal of less perfect screens within it and, so far as is possible, as one terminal of generators and detectors. There is then no need that the screen enclosing the entire apparatus should be connected to The Earth; indeed, in a high-voltage laboratory that may be positively undesirable or impossible; one man's earth may have to be another man's II.'I'. terminal. Accordingly, even if *earth* could always be identified with a common terminal in the sense just described, it would still be necessary to recognize that "earth" does not always mean The Earth, and, more generally, that there is no single body that can always be taken as earth.

There is another statement about "earth" made or implied in conventional text-books, especially in the description of the classical experiments of electrostatics, on which the whole subject is supposed to be based. It is that, if any charged body is connected to earth, it will be wholly discharged. According to § 13, this will be true only if "earth" means a screen surrounding the body completely or practically completely. If that is what "earth" means, earth cannot possibly be a single body everywhere at the same potential; for, if there are two charged capacitors in series, the screen of one is necessarily the plate of the other. Nevertheless, the use of "earth" in this sense may have some significance in connection with the said classical experiments. For if they were conducted in the open, some of them would fail owing to the Earth's atmospheric gradient; if they are to succeed, they must be conducted within a screen, which will have to be the more complete, the more accurate the real experiments. The classical experiments are so inaccurate that the very incomplete screen formed by the partially conducting walls, ceiling and floor of the building in which they are conducted may suffice. When a charged body is discharged by touching it with the finger or with a wire connected to the water pipes, that may be because such contact provides a sufficiently conducting path between the body and the imperfect screen by which it is surrounded.

But another explanation is possible. The charge in a capacitor will be greatly reduced if its terminals are connected to an uncharged capacitor of much greater capacitance. It is possible that a charged body is discharged by the finger of the observer, not merely because he provides a conducting path to the screen, but because his body forms with all those conductors, with which the charged body has mutual capacitance, a capacitor of much greater capacitance; or, more simply, because his body is much larger than the body he discharges.

This, we suggest, is the basis of the curious statement to be found in some well known text-books (it appears to have grown commoner during the last generation) that 'The Earth acts as an "earth"', because its capacitance is much greater than that of any body concerned in terrestrial experiments. But, though it may provide the basis, it provides no justification. For, firstly, the statement is not actually true; the capacitance of 'The Earth', as ordinarily calculated, is no greater than that of certain electrolytic condensers that are often used in laboratory experiments. Secondly, and more important, the capacitance of 'The Earth' cannot possibly have any relevance to any terrestrial experiment. The facts that give significance to the conception of the capacitance of a single body (not of a pair of associated bodies) are these: (they are predicted by the field equations and confirmed by experiment.) If in a closed capacitor the distance between the plate and the screen is increased without limit, the capacitance of the capacitor tends to a lower limit; the capacitance of a single body means this lower limit towards which the capacitance of a closed capacitor tends. But since there is no accessible conductor whose distance from 'The Earth' is large compared with 'The Earth's' dimensions, this limit can have no relevance to experiments conducted on 'The Earth'.

§ 15 POINT CHARGES

Point charges play a very important part in classical theory—much more important than that which they play in experimental physics. It seems therefore important to show that the propositions that we have set forth lead to the chief propositions about point charges. (These are Coulomb's law, and

$$F = e\mathcal{E} = e \cdot dv/dq, \quad \dots\dots(15.1)$$

where F is the force in the direction of the coordinate of q acting on a point charge e in a field of potential v (§ 5).

Let us consider then the force between two charged conductors, which should lead us to Coulomb's Law. For reasons already given we must include in our system an earth, i.e. a third conductor surrounding the two small ones (figure 4a); but in order to simplify the problem we shall assume that the third conductor is so large that the distance r between 1 and 2 is very small compared with their distances from 3, but large compared with the linear dimensions of 1 and 2. Unless the charges on 1 and 2 are equal and opposite, there must be a charge equal to their difference on the inner surface of 3; there may also be charges on its outer surface, but they do not concern us. We have from (11.1) with an obvious notation

$$W_V = \frac{1}{2}(C_{12}V_{12}^2 + C_{23}V_{23}^2 + C_{13}V_{31}^2) \quad \dots\dots(15.2)$$

$$\text{with} \quad Q_1 = C_{12}V_{12} + C_{13}V_{13}, \quad \dots\dots(15.3)$$

$$Q_2 = C_{12}V_{21} + C_{23}V_{23}, \quad \dots\dots(15.4)$$

$$V_{23} = V_{21} + V_{13}. \quad \dots\dots(15.5)$$

Substituting for the V s in (15.2) from (15.3, 4, 5), we find after some reduction

$$2W_Q(C_{12}C_{23} + C_{23}C_{13} + C_{13}C_{12}) = Q_1^2(C_{12} + C_{23}) + Q_2^2(C_{12} + C_{13}) + 2Q_1Q_2C_{12} \dots (15.6)$$

In case (a), if the bodies 1 and 2 are small compared with their distances from the screen and from each other, C_{12} is small compared with each of C_{13} , C_{23} . It follows then from (15.6) that, if F_r is the force tending to increase the distance r between 1, 2, then

$$F_r = -dW_Q/dr = -\left\{ \frac{1}{2}Q_1^2 \frac{d}{dr} \left(\frac{1}{C_{13}} \right) + \frac{1}{2}Q_2^2 \frac{d}{dr} \left(\frac{1}{C_{23}} \right) + Q_1Q_2 \frac{d}{dr} \left(\frac{C_{12}}{C_{13}C_{23}} \right) \right\} \dots (15.7)$$

Now, according to a well known result (see e.g. Jeans, *Electricity and Magnetism*, § 116), if a , b are the radii of spheres to which the bodies 1, 2 approximate, and if $ab \ll r^2$

$$C_{13} = \kappa a; \quad C_{23} = \kappa b; \quad C_{12} = \kappa ab/r, \dots (15.8)$$

where κ is the permittivity of the medium surrounding the bodies. Hence the terms of (15.7) in Q_1^2 and Q_2^2 vanish, and we have

$$F_r = -Q_1Q_2 \cdot d/dr(\kappa ab/r) = \dots (15.9)$$

$$= +Q_1Q_2/\kappa r^2, \dots (15.10)$$

which is Coulomb's Law. Thus, although the law is not strictly true for small charged conductors, as is obvious from (15.7), it is an approximation that becomes true in the limiting case in which the charged bodies become infinitely small.

Consider now the force on a single small charged body 2 in the field between two large bodies 1 and 3 (figure 4b). This is another case of the general problem of three charged bodies, and (15.6) again holds; but now C_{12} , C_{23} are both small compared with C_{13} , as in (say) the oil-drop experiment. So we have

$$F_q = -dW_Q/dq = -\frac{1}{2}Q_1^2 \frac{d}{dq} \left(\frac{1}{C_{13}} \right) - Q_2^2 \frac{d}{dq} \left(\frac{1}{C_{12} + C_{23}} \right) - Q_1Q_2 \frac{d}{dq} \left(\frac{1}{C_{13}} \cdot \frac{C_{12}}{C_{12} + C_{23}} \right), \dots (15.12)$$

where q determines the position of the small conductor relative to the large ones, whose relative position is constant.

The term in Q_1^2 represents a force, important in some circumstances, tending to bring 2, even if it is uncharged, into the region where the field due to the charges on 1 and 3 is strongest. Since e in (15.1) is our Q_2 , there is no corresponding term in (15.1) independent of Q_2 . The term in Q_2^2 represents (in conventional language) the "image force" between Q_2 and the charges that it "induces" in the large conductors. If v is independent of Q_2 , as the form of (15.1) suggests, there is again no corresponding term in (15.1); but one can be introduced by including in v a part due to the "induced charges"; a part due to the charge Q_2 on 2 must not be introduced, for such a part would not give rise to a force on 2. Accordingly (15.1) in its natural interpretation is at best an approximation valid only when the terms of (15.7) in Q_1^2 and Q_2^2 are negligible. That in Q_1^2 is negligible, if the field

due to the charges on 1 and 3 is sufficiently uniform and the dimensions of 2 sufficiently small. The term in Q_2^2 is negligible, if Q_2/Q_1 is sufficiently small and if also 2 is sufficiently far from 1 and 3; for, when it is far from 1 and 3, dC_{12}/dq and dC_{23}/dq tend to be equal and opposite.

If these conditions are fulfilled (15.7) reduces to

$$F_q = -Q_1 Q_2 \cdot \frac{d}{dq} \left(\frac{1}{C_{13}} \cdot \frac{C_{12}}{C_{12} + C_{23}} \right); \quad \dots\dots(15.13)$$

and this is the equation that really has to be compared with (15.1). In that equation v is the potential relative to the screen 3 of the point occupied by 2, when 1 is at its actual potential V_{13} . This may be identified with the voltage that would be established between 3 and a small uncharged conductor in the position of 2, if the voltage between 1 and 3 were increased from 0 to V_{13} . The small conductor, being insulated, cannot become charged during the process; and we have from (15.4)

$$0 = C_{12}V_{21} + C_{23}V_{23} + C_{12}(v - V_{13}) + C_{23}v$$

or

$$v = V_{13} \cdot C_{12} / (C_{12} + C_{23}). \quad \dots\dots(15.14)$$

Hence (15.13) becomes

$$F_q = -Q_1 Q_2 \cdot \frac{d}{dq} \left(\frac{v}{C_{13}V_{13}} \right) \quad \dots\dots(15.15)$$

or from (15.3),

$$= -Q_2 \cdot \frac{d}{dq} \left(v \cdot \frac{Q_1}{Q_1 - C_{12}V_{12}} \right), \quad \dots\dots(15.16)$$

(15.16) is identical with (15.1) only if $C_{12}V_{12} = Q_1$. This does not follow immediately from (15.3), (15.4) together with $Q_2 \ll Q_1$ and $(C_{12}, C_{23}) \ll C_{13}$; for the two terms in (15.4) may be of opposite sign. But it can hardly be doubted that there are conditions in which (15.16) approaches (15.1) as a limit as Q_2 approaches zero. We leave to the mathematician the task of deciding exactly what those conditions are; they will certainly involve the interpretation given to v in (15.1). Further, a correction, which will certainly involve terms in Q_2/Q_1 only, has to be made for the fact that, while (15.1) relates to a state in which V_{13} remains constant during the motion of Q_2 , we have calculated the force on the assumption that Q_1 remains constant during that motion. It is sufficient for our purpose to note that (15.1) can be deduced by the argument that we have presented only as a highly specialized limit.

The complexity of the argument necessary to deduce the very simple classical formula (15.1) may suggest the criticism that, though our exposition of electrostatics may be ultimately sounder than that of classical theory (at least if that theory is supposed to have any relation to experiment), the classical treatment is always to be preferred in dealing with any practical problem. We reject any such suggestion. The physicist who applies the simple formula (15.1) to any real experimental system must always make due allowance for the difference between his actual system and the ideal system for which the formula is strictly valid. The simplicity of (15.1) is obtained only at the cost of complexity of interpretation in terms of measurable quantities, and our methods remove this complexity. For general problems in experimental physics we suggest that our methods are far more perspicuous physically, and no more difficult algebraically, than classical

methods. This is particularly true of the problems (§§ 10–12) arising in “systems of conductors”; actual experience has convinced us that the treatment of such problems in terms of the equivalent capacitance network has great advantages.

§ 16. CHARGES ON INSULATORS

It has been assumed in the two preceding sections, explicitly or tacitly, that charged bodies are always conductors. This is not true. Of the bodies which, when introduced into a Faraday cage produce a non-zero δV_{12} , some are insulators. Such bodies cannot form a plate of a capacitor, cannot be charged or discharged by a current circuit, and have no measurable voltage to other bodies; accordingly many of the propositions of those sections have no application to them. Nevertheless they have a determinate charge, because they satisfy the criterion of § 14, namely that n equivalent bodies, introduced together into the cage, produce n times the effect of one of them. Further, if we suppose that a conductor and an insulator, having the same charge according to the Faraday cage, differ only in the distribution of the charge, all propositions concerning point charges must be equally true of both of them; for in the limit when the dimensions of the body are zero, the distribution of the charge in the body cannot matter. Accordingly (15.1) can be applied to the measurement of the charge on an oil drop, by observing the force on it in a uniform field, although the drop may be an insulator; or to the deflection of a cathode ray in an electric field, without making an assumption about the constitution of an electron, except that it is small and charged.

But the distribution of charge on an insulator enters into other problems. They can be solved in a manner that we shall not discuss in detail by means of a proposition about the relation between the field v (§ 6) about an insulator and the distribution of charge in and over it. The conventional proposition is that if σ is the surface density of the charge and ρ the volume density, then v must satisfy the differential equations

$$4\pi\sigma = - \left(\frac{dv}{dn_1} + \frac{dv}{dn_2} \right), \quad \dots\dots(16.1)$$

where n_1, n_2 are the normals to the surface on opposite sides of it

$$4\pi\rho = - \left\{ \frac{d}{dx} \left(\kappa \frac{dv}{dx} \right) + \frac{d}{dy} \left(\kappa \frac{dv}{dy} \right) + \frac{d}{dz} \left(\kappa \frac{dv}{dz} \right) \right\}. \quad \dots\dots(16.2)$$

The direct evidence for these propositions is very slight. Some evidence in favour of (16.1) can be obtained by means of “proof planes”, i.e. small conductors applied to the surface of the insulator and then transferred to a Faraday cage, if it is assumed that the proof plane takes up the surface density of the surface to which it is applied. (It may be noted incidentally that, by means of proof planes, evidence can be obtained for the presence of the charges $-Q_1'$ and $+Q_1'$ on the inner and outer surfaces of the intermediate plate of the two capacitors in series discussed in § 5.) Volume charge can be detected only by breaking the body and transferring the fragment to the cage. Actually volume charges in solids are at present of little experimental importance. The force between two bodies, one at least of which is an insulator, arising from their charges, can be obtained by means of a proposition

that might have been stated earlier. It can be shown by purely mathematical reasoning that (11.1), (11.2) are equivalent to

$$W = \frac{1}{8\pi} \int \left\{ \left(\frac{dv}{dx} \right)^2 + \left(\frac{dv}{dy} \right)^2 + \left(\frac{dv}{dz} \right)^2 \right\} dx dy dz. \dots (16.3)$$

It is assumed that this formula for the energy holds even when some or all of the charges reside on insulators. The results deduced from this assumption, together with (16.1), (16.2), apparently agree with such experiments as can be made.

Finally, it might appear that, if the dielectric forming part of a capacitor were charged, the rule for calculating capacitance given in §4 would no longer hold. Experiment, in accordance with deduction from (16.1), (16.2), shows that such charges, though they modify the values of ϵ , do not modify the capacitance.

§ 17. INDUCED CHARGE

In the conventional exposition of electrostatics a magnitude called *induced charge* is introduced, and considerable space is devoted to a discussion how far "induced" and "real" charges have the same properties. The expression *induced charge* is often useful in drawing attention to the fact that, when an insulated conductor is introduced into an electric field, the distribution of the charge on it, as measured by ϕ , changes, while its total charge, measured by $\int Idt$, does not.

But it appears to us that attempts to give the expression definite quantitative significance introduce confusion rather than clarity.

REFERENCES.

- CAMPBELL, N. R. and HARTSHORN, L., 1946, *Proc. Phys. Soc.*, **58**, 634.
 HARTSHORN, L., 1940, *Radio-Frequency Measurements by Bridge and Resonance Methods* (London: Chapman and Hall).
 ROSA and DORSEY, 1907, *Bull. Bur. Stand.*, **3**, 43

Measurement of Absolute Humidity in Extremely Dry Air

By A. W. BREWER, B. CWILONG AND G. M. B. DOBSON,

Oxford

MS. received 29 May 1947

ABSTRACT. To meet an urgent demand for a hygrometer capable of use at all heights in the atmosphere, the dew-point hygrometer, which is shown to have many advantages, has been developed primarily for use in aircraft to measure dew-points, or rather frost-points, down to -90°C .

It is necessary that the instrument should operate at the lowest possible frost-points as it has been discovered that the air in the stratosphere is very dry. Laboratory studies of the deposition of water and ice from the vapour at low temperature are described.

Below -90°C , it is not possible to operate a frost-point hygrometer because the deposit is in the form of an invisible glassy layer, but the instrument gives correct results at temperatures close to this limit.

Details are given of the construction of different forms of hand-operated hygrometers, and work is now going on to develop a fully automatic frost-point hygrometer.

§1. INTRODUCTION

METEOROLOGISTS have long felt the need for an instrument capable of measuring the water content of the upper air at altitudes where the low temperatures preclude the use of the usual methods. The method of measuring the water content of air, or any other gas, down to a concentration of 0.4 mg./m^3 which is described here, has been developed primarily for use in the upper air from aircraft, but it is quite suitable for experimental use for any other purpose.

The interest in the humidity of the upper air is due to the fact that the water content at any level usually gives an indication of the past history of the air at that level. For this particular problem a hygrometer is needed which retains its accuracy at very low relative humidities, for instance one which can distinguish between 1% R.H. and 2% R.H. in air at a very low temperature. To do this the hygrometer must be capable of measuring small quantities of water vapour with an accuracy not previously attainable.

The average temperature of the stratosphere in temperate latitudes is about -56°C . (This is the temperature of the stratosphere in the ICAN standard atmosphere), and air saturated at this temperature would have a water vapour content of about 18 mg./m^3 . When this work was begun it was expected that this concentration, or a value only slightly below it, would be the lowest that would have to be measured. When, however, flights were made into the stratosphere it was found that the relative humidity there was of the order of only 1 or 2% , corresponding to saturation temperatures down to -80°C . In order, therefore, to measure the water content of the stratosphere it has been necessary to develop a hygrometer capable of measuring water contents of less than 1 mg./m^3 .

Most of the hygrometers ordinarily used in meteorological work are quite unsuitable for use at low water contents and low temperatures. At low temperatures the usual wet and dry bulb psychrometer has a vanishingly small temperature difference even for 0% R.H., and also the constant in the psychrometric formula is in doubt (Griffiths and Awbery 1935). The hair hygrometer, the gold-beater's skin hygrometer, and other hygrometers employing organic elements sensitive to relative humidity have a large lag which increases rapidly at low temperatures so that they cannot be used. Electrical conduction hygrometers, such as the Gregory humidimeter, have a large temperature coefficient, which seriously interferes with their use at low temperatures, while in addition the lag increases to an undesirably large value. For further discussion of these difficulties, reference may be made to Ebert (1937) and Simons (1936), who discuss the difficulties of low-temperature hygrometry. Gluckauf (1945) discusses the effect of lag on the hair hygrometers in meteorograph measurements and Diamond *et al.* (1940) discuss the properties at low temperatures of the electrical hygrometric element of the U.S. radiosonde. Dymond (1947) gives a similar discussion for the gold-beater-skin element of the British, Kew pattern, radiosonde; he also quotes results obtained by Harrison and Brewer (1944) of comparison ascents in which an aircraft, equipped with a hygrometer of the type described in §3 of this paper, circled an ascending radiosonde balloon.

The only methods which appear at all practicable are the dew (or rather hoar frost) point hygrometer, and a radiation absorption method. The latter method could only be operated in the 6.7μ absorption band on account of the

very small quantities of water involved and, even so, would require very long absorption paths.

The optical absorption hygrometer was carefully considered, but it was decided that it would present difficulties at very low vapour pressures and also much basic work on the absorption by water vapour of infra-red radiation at low pressures and temperatures would be required. The radiation method is most likely to be successful if the absorption of solar radiation in the $6.7\ \mu$ band is measured, so as to give the total water absorption above the observer.

The dew point hygrometer on the other hand has been used by several workers down to about -30°C . without giving special difficulties and it seemed hopeful that it could be developed to operate at much lower temperatures. The instruments described here have been used regularly at frost-points down to -80°C ., but there is a natural limit to their operation at about -90°C . which will be discussed fully in §2. Since throughout this work the water contents measured have almost invariably been below the saturation vapour pressure at 0°C ., the instruments are usually referred to as frost-point hygrometers, particularly as they are more sensitive to a hoar-frost deposit than a dew deposit, and this name will be used below.

The frost-point hygrometer has many advantages. The water content is measured upon a nearly logarithmic scale quite independent of the actual air temperature, and this makes for high discrimination at low relative humidities, provided that the water content is not so very low as to correspond to a frost-point below -90°C . For meteorological purposes also the observations can be used directly by considering the frost point, usually plotted with the air temperature, upon a thermodynamic diagram. For use in an aircraft, it is also very helpful to be able to bring the air into the aircraft in a duct and measure the frost-point in a convenient apparatus, for, provided that the walls of the duct do not give off or absorb water vapour, then there is no change in the dew- or frost-point. A hygrometer which measures the relative humidity has, by comparison, many difficulties. Another very real practical advantage is its basic stability. For calibration it is only necessary to calibrate the thermometer which measures the thimble temperature. The thermometry involved is very simple, and re-calibration is not normally required even after long periods of either use or disuse. Another advantage is that supersaturation which is not uncommon in the upper air is shown quite simply, as the frost-point is then measured as being higher than the air temperature, which is measured separately.

To illustrate the practical use of the frost-point hygrometers described below the results of two ascents are shown in figure 1 and figure 2. Figure 1 is an ascent to 12 km. made during unsettled weather and shows moist air in the troposphere and the usual very dry air in the stratosphere. Figure 2 is an ascent to 4 km. made in an anticyclone with observations at frequent intervals to show the details of the distribution of water vapour in the stable layer between 1 and 2 km. Both these ascents were made with eye-observation instruments of the type described in §3. It is hoped to discuss the meteorological aspects of these and other ascents elsewhere.

§2. DEPOSITION OF WATER VAPOUR AT VERY LOW TEMPERATURES

In addition to the obvious difficulties involved in observing very small amounts of dew or frost, a number of practical difficulties are found when the

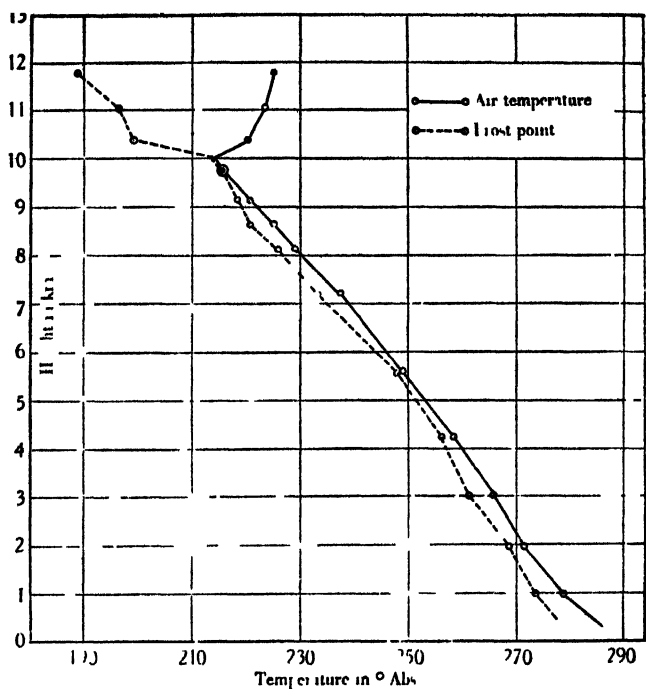


Figure 1 Frost-points and air temperatures observed on Boscombe Down, 30 May 1945.
1300 CMT The frost-point at the top is the lowest yet observed
(Reproduced, by kind permission, from the 'Proceedings of the Royal Society')

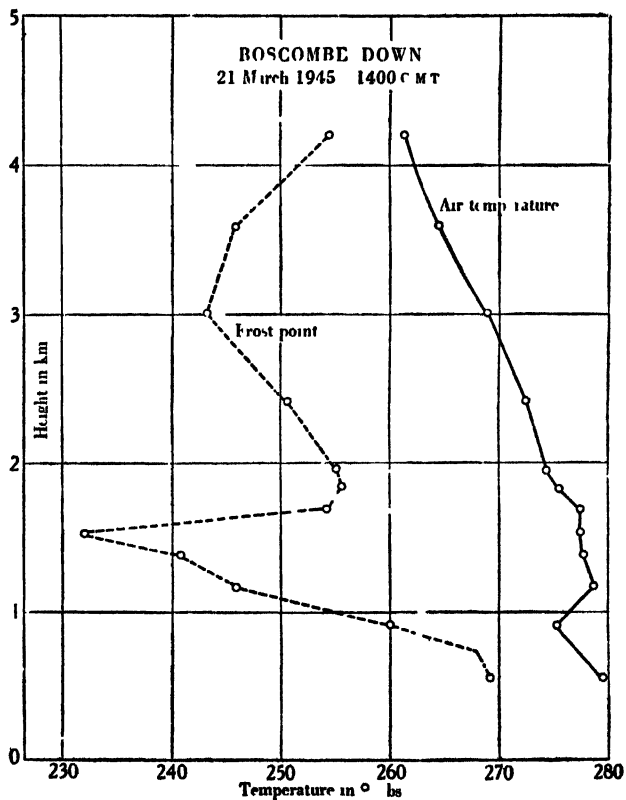


Figure 2 Frost-points and air temperatures observed in an anticyclonic inversion 21 March 1945.
(Reproduced, by kind permission, from the 'Proceedings of the Royal Society')

humidity of very dry air is measured by the frost-point type of hygrometer. Some of these have a wider interest. They may be summarized as follows:—

- (a) Uncertainty whether the deposit is dew or hoar frost.
- (b) Hoar frost is not necessarily deposited when the thimble temperature is below the frost-point.
- (c) Slowness of deposition and evaporation at low temperatures.
- (d) Gradual change in the character of the deposit with falling frost-point, from large white ice crystals, through a rather uniform white deposit, to a bluish bloom, and finally to an invisible glassy film.

We shall consider each of these difficulties in turn.

(a) When water vapour is deposited on a solid surface below 0° c. it does not follow that ice crystals will be formed. If the surface is polished and clean, dew will generally be deposited down to temperatures as low as -40° c., while by taking special precautions it is possible to form dew on a solid surface at any temperature down to -100° c. If the surface is not very clean ice may be formed at temperatures well above -40° c.

Table 1. Saturation vapour pressures on water and ice
(Brunt, *Physical and Dynamical Meteorology*)

Temperature (°c.)	270	260	250	240	230	220	210	200
Sat. v.p. (mb.) $\left\{ \begin{array}{l} \text{over water} \\ \text{over ice} \end{array} \right.$	4.89 4.76	2.25 1.98	0.97 0.77	0.38 0.27	0.14 0.09	0.045 0.027	0.013 0.007	0.0035 0.0017
Ratio $\frac{\text{v.p. ice}}{\text{v.p. water}} > 100$	97	88	79	71	65	60	54	49

If ice is deposited at, say, -5° c., a few big crystals are formed. On the other hand if dew is deposited at the same temperature, the surface is uniformly covered with very fine droplets. There is, therefore, no difficulty in distinguishing between dew and hoar frost at such temperatures. At lower temperatures, say about -30° c., to -40° c., both dew and hoar frost may cover the surface with small water drops or fine crystals and it is, at times, very difficult even for an experienced observer to be certain whether a deposit is dew or hoar frost. Table 1 shows the saturated vapour pressure over water and over ice at different temperatures, and also the relative humidity, with respect to water, of air saturated with respect to ice. It will be seen from these figures that it is important to be sure of the character of the deposit. For this reason much time was spent in trying to find a surface, suitable for the thimble of a frost-point hygrometer, on which the deposit would always be either water or ice. Unfortunately, although clean surfaces were found which had such properties when freshly made, they tended to lose these properties when they became old. (Further details of this work will be published in another paper and it will only be summarized below.)

On a freshly made clean zinc surface, formed by deposition from the vapour in a vacuum, ice crystals are normally formed at all temperatures below 0° c. A surface of freshly cleaved mica is the extreme opposite to zinc and it seems impossible to form ice crystals initially on such a surface, a uniform film of water being always formed at first at all temperatures down to -100° c., though it may freeze later. Both these surfaces soon "age" and then they act more like other common surfaces, dew being formed above about -30° c. to -40° c. and hoar frost at lower temperatures. A cadmium surface is similar to zinc, while a

polished aluminium surface is more like mica. All polished metal surfaces have a greater tendency to have dew deposited on them than crystalline surfaces of the same metal. When measuring saturation temperatures at which either dew or hoar frost may be formed and are difficult to distinguish, we find it best to cool the thimble sufficiently to freeze any deposit of dew and then to allow it to warm up until this ice deposit has nearly, but not quite, evaporated.

Various experiments were made to try to "seed" the thimble surface so that hoar frost would be formed instead of dew, but without success. Crystalline quartz dust appeared to be a likely substance since both quartz and ice form hexagonal crystals. It was, however, quite without effect, and even if quartz crystals were dropped into supercooled water, they had no more tendency to cause crystallization than glass dust. It has been suggested that quartz dust in the atmosphere would form natural nuclei on which water vapour would sublime at low temperatures to form ice crystals, but apparently this is not the case.

It is of some interest to note that dew can be formed on most surfaces at any temperature above about -100°C . if the proper precautions are taken—the dew, of course, often freezes rapidly after being formed. Thus, if the thimble of a dew-point hygrometer be cooled to a low temperature in a current of air which is so dry that no deposit is formed and then a puff of damp air is admitted, dew will be formed on the thimble. Even if the thimble has a number of ice crystals scattered over it dew will be formed on the surface between the crystals.

A further point of interest is that two supercooled water droplets may grow on the thimble until they touch and coalesce and they will still remain liquid. Thus it appears probable that supercooled cloud and rain droplets which may collide in the air will not necessarily freeze as has sometimes been suggested.

(b) When the thimble of a dew-point hygrometer is cooled and hoar frost is deposited, it is almost always found that the hoar frost is not deposited immediately the thimble temperature falls below the frost-point, but that it has to be cooled nearly to the dew-point before any deposit is formed. The temperature of the thimble when the first trace of frost appears is therefore not a good indication of the frost-point even if the thimble be cooled very slowly. Nor is the mean of the temperatures when a deposit is first seen on cooling and last seen on warming an accurate measure of either dew- or frost-point. One must find the temperature of the thimble when the deposit is neither growing nor evaporating.

(c) At the lower temperatures, say -70°C . and -80°C ., the amount of water vapour in the air is so small and the vapour pressure is so low, that both the growth and evaporation of the deposit are very slow. If an eye-observation instrument is used, it is necessary to wait at least a minute with the thimble held at a constant temperature to be sure whether the deposit is just growing or just evaporating. This makes the measurement very tedious and the photo-electric deposit indicator has the great advantage that a small change in the amount of the deposit can be quickly seen.

(d) As drier and drier air is used in a frost-point hygrometer the nature of the deposit gradually changes. For frost-points above about -30°C . the deposit, if ice, consists of relatively large white crystals which can easily be seen with a magnifying lens. At lower temperatures the deposit consists of a uniform layer of fine crystals, and below a temperature of about -70°C . these crystals are so fine that a thin deposit has a definite blue colour, while at about -80°C .

the deposit is often difficult to see and appears as little more than a faint bloom on the thimble. At still lower temperatures no visible deposit is formed at all, and the water vapour is apparently condensed as a uniform glassy layer. Even if the thimble be first coated with a thin layer of hoar frost and then cooled to say -120°C . in air of frost-point about -100°C ., the deposit does not appear to grow, even if it is measured by the very sensitive photoelectric deposit indicator described later. That a layer of glassy ice is actually formed in the above case when the thimble is cooled below the frost-point in very dry air is shown by the following experiment. The thimble is first cooled to, say, -120°C . in a stream of extremely dry air (obtained by passing air through a spiral, cooled in liquid air), then air of frost-point about -70°C . is passed for a minute or two, during which time there is no visible change in the thimble surface; the extremely dry air is then again passed and the thimble slowly warmed. As the thimble warms up to a temperature of -90°C . to -80°C . a deposit appears to form although it is in air of frost-point far below this temperature, and if held at this temperature the deposit slowly evaporates again. There seems no doubt that an invisible deposit of glassy ice is deposited at the very low temperature and that as it warms up to about -90°C . it begins to turn into crystalline ice. The formation of glassy ice at low temperatures was confirmed by an experiment made in a vacuum. A molecular beam of water vapour was directed against a cooled polished gold surface which was illuminated with monochromatic light. When the temperature of the surface was above about -90°C . a visible deposit was formed as usual and no optical phenomena were observed. When the temperature of the surface was below -90°C . Newton's rings were observed, and these moved as the layer of glassy ice increased in thickness. If, after deposition at very low temperatures, the surface was slowly warmed, the deposit changed on passing a temperature of about -90°C . to a translucent layer and the rings disappeared.

This phenomenon of the formation of glassy ice is unfortunate since it appears to set a lower limit to the frost-point which can be measured with such an instrument. Though this limiting temperature is low, about -90°C ., frost-points have been observed in the stratosphere almost down to the limiting values, and we should certainly like to be able to measure still drier air.

§ 3 DESCRIPTION AND USE OF THE HYGROMETERS

The essential purpose of the instrument is to ventilate with the air under measurement a surface covered with dew or frost, and to find the temperature of the surface at which the dew or frost is in equilibrium with the air. The prime requirements of the instrument are therefore (a) a properly ventilated surface, the temperature of which can be easily and smoothly controlled and accurately measured, and (b) satisfactory arrangements to determine whether the deposit of dew or frost is changing. In the visual hygrometer any change in the amount of the deposit is judged visually, and the eye is given as much assistance as possible.

The test surface is formed by the top surface of the thimble, seen in the diagram (figure 3). It is a relatively massive piece of pure aluminium and is, for special reasons, anodized and dyed black. Its thermal inertia prevents too rapid changes in temperature. Its temperature is adjusted by balancing a controlled cooling against natural heating, which is provided by the heat of the air stream ventilating the top of the thimble and by thermal conduction from

other parts of the instrument. The conduction is kept at a low value by supporting the thimble on a sheet of paxolin $\frac{1}{8}$ " thick, and surrounding the skirt with a sleeve which prevents heating of the skirt by convection. An electrical

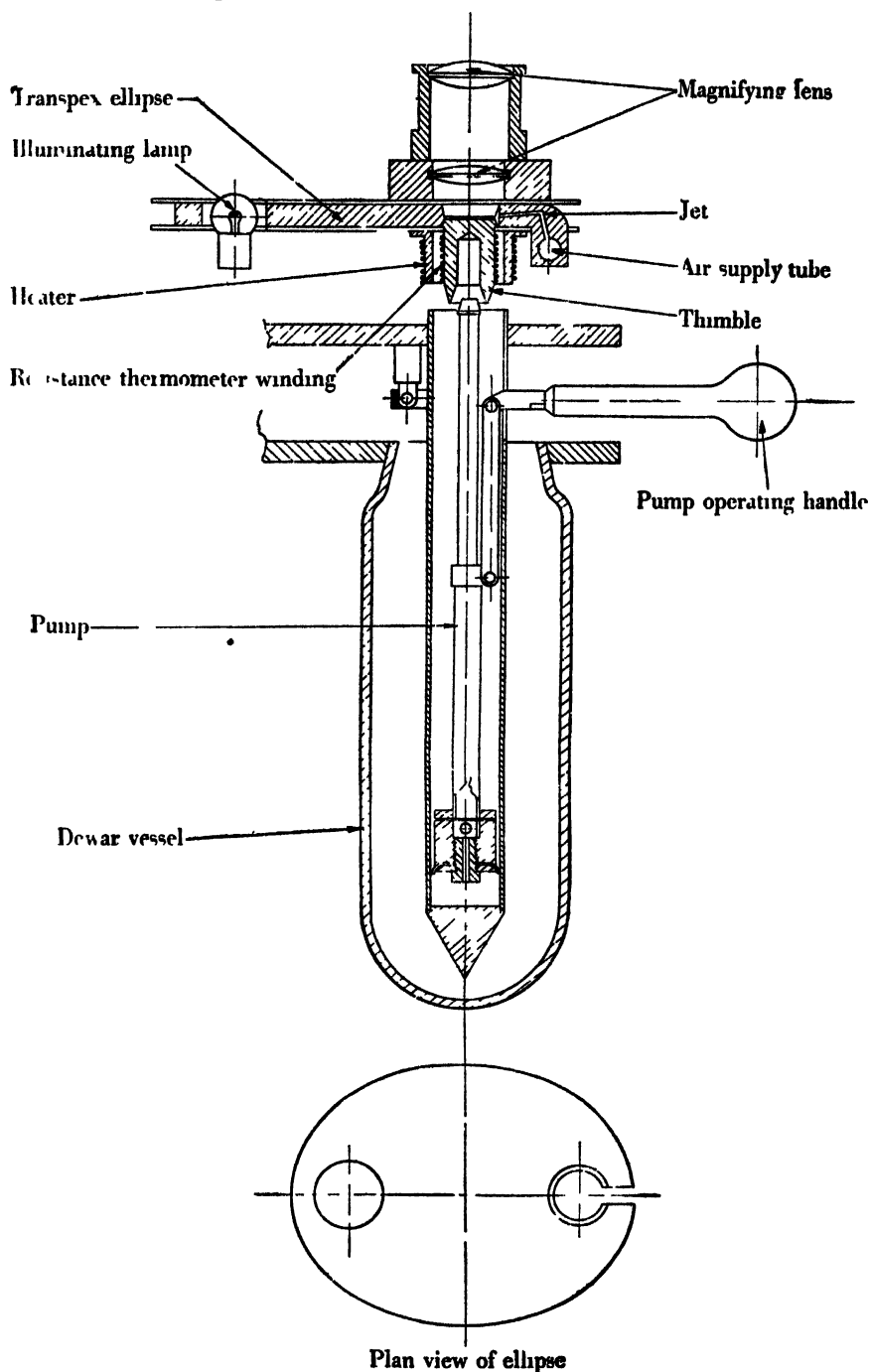


Figure 3. Diagram showing the principle of the eye-observation frost-point hygrometer. (Reproduced, by kind permission, from the 'Proceedings of the Royal Society'.)

heater is wound on this sleeve for use when natural heating is insufficient. Its use is discouraged by introducing a switch for the electric supply which must be held on during the period that additional heating is required.

The cooling is carried out by forcing a coolant as a jet against the hollow base of the thimble. The pump used for this is a small, hand-operated piston pump carefully designed to give the minimum conduction from the coolant to the top of the pump, and also to give reliable operation under the conditions of use. Excessive conduction to the upper parts of the pump is undesirable both because of the loss of coolant which results and also on account of the large amounts of hoar frost which accumulate on the exposed parts of the pump. The pump is held by a flat bayonet joint in a flange at the top, which attaches it to a part of the main structure of the instrument entirely separate from the thimble. The object of this design is to avoid conduction from the top of the pump to the thimble. The coolant used in the hygrometer shown in figure 3 is a mixture of alcohol, acetone or petrol with solid carbon dioxide. This cools the petrol, alcohol or acetone, which is contained in the pump and which is pumped up against the thimble and then drains back into the pump where it is again cooled ready for recirculation. This is necessary on account of the difficulty of pumping a sludge of solid carbon dioxide.

The hygrometer shown in figure 4 is a modification of that of figure 3, re-designed for use in aircraft with a pressure cabin. It is essentially intended for use at very high altitudes where carbon dioxide cooling is inadequate. The refrigerant used in this case is liquid air, and the liquid air itself is circulated by the pump. For this reason a supply valve is fitted at the bottom of the pump and there is an outlet hole in the side. The liquid air pump must operate at atmospheric pressures down to 150 mm. of mercury, and since the liquid air is boiling, considerable difficulty is experienced in getting the liquid air into the pump during the upward stroke of the piston. For this reason the liquid air inlet valve at the bottom of the pump must be easily operated and of generous size.

The thimble is of pure aluminium and is of such proportions that, with its relatively high thermal conductivity, any significant temperature differences between different parts cannot be maintained except for a very short time. It is thus possible to measure the temperature of the thimble top by a resistance thermometer wound on the skirt. For use in aircraft, a resistance thermometer is preferable to a thermocouple thermometer since very much greater power is obtainable from it to operate an indicator, which can therefore be made more robust. Also the resistance thermometer permits measurement of the thimble temperature, and hence the dew-point or frost-point, directly, and this in practice is another advantage. A very satisfactory thermometer is obtained by cutting a very fine thread, 100 turns per inch or finer, on the skirt of the thimble. Two fine skew holes are also drilled in the walls of the skirt and the thimble is then anodized, thus providing an excellent insulating coating upon the thread and in the fine holes. Leads of enamelled copper wire of moderate gauge are then cemented in the skew holes with a bakelite cement in such a way that a tip of the copper wire just protrudes and to this the resistance wire is soldered. The resistance thermometer wire, of $50\ \mu$ diameter bare platinum, is soldered first to the projecting tip of one copper lead and then wound down the thread on the skirt to the other lead, to which it is soldered. The wire lies at the bottom

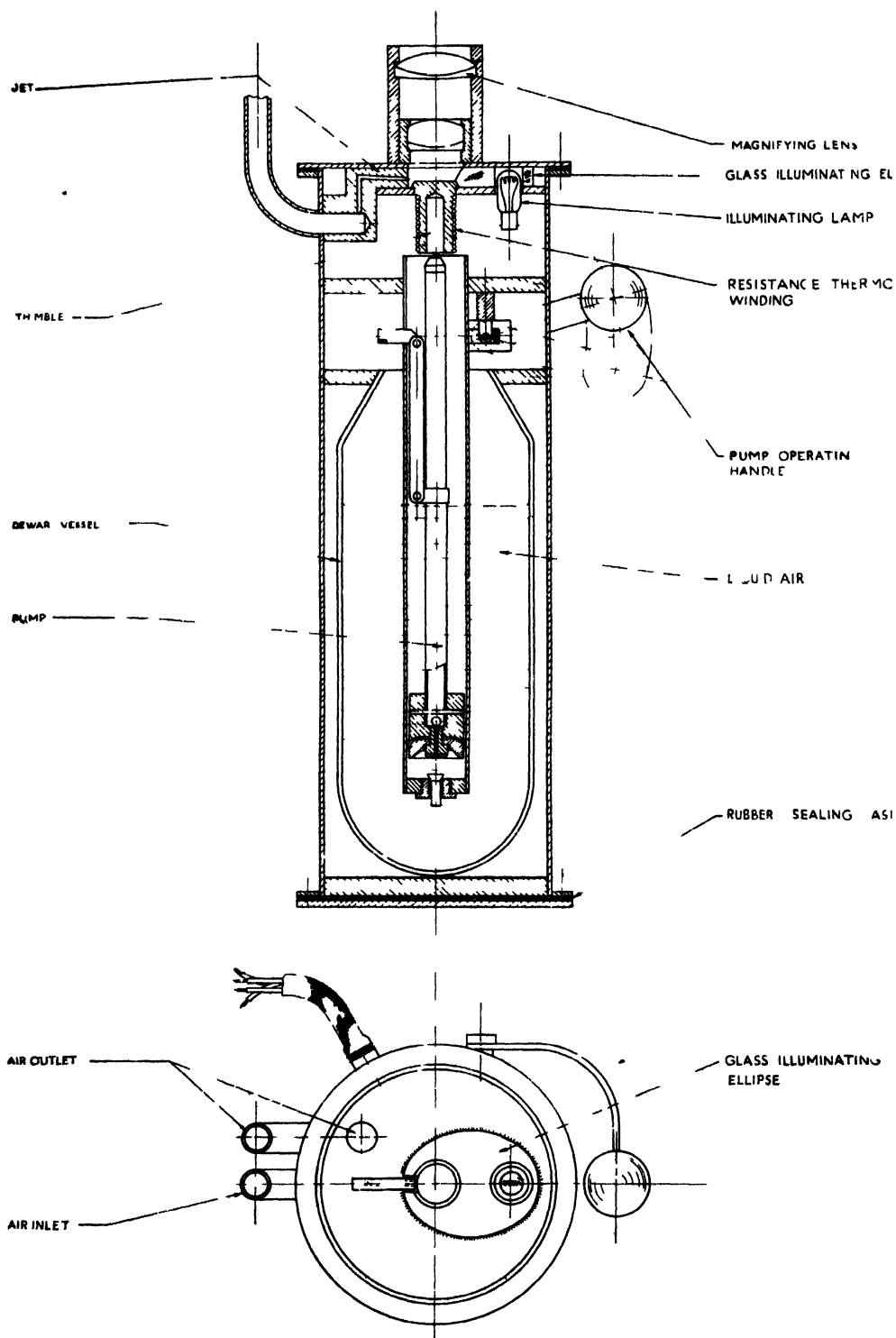
of the groove of the thread so that the bare wire is brought into a very intimate contact with the thimble and a thermometer is formed with no detectable lag. To protect the winding and improve the contact the sides of the thimble are given a coat of bakelite varnish which is baked on. This technique of winding a fine bare wire in an anodized thread cut upon an aluminium surface has proved useful in several ways; it is, for example, used in the heater which surrounds the thimble.

The top surface of the thimble, the test surface, is ventilated by a stream of air from the jet which is a 1 mm. diameter hole pointing downward at an angle of 10° and directed at the centre of the test surface. Air from outside the aircraft is supplied to the instrument through a short length of clean $\frac{1}{4}$ " copper tube. A rapid stream of air passes constantly through the tube and a small proportion is forced to flow through the hygrometer jet by reducing the outlet of the tube to $3/16$ ". The surplus air which flows through the $3/16$ " hole serves to keep all the supply pipes well scavenged and is utilized to provide a forced ventilation to the illuminating lamp.

It is not satisfactory to use daylight as a source of illumination for viewing the deposit because in an aircraft the illumination is liable to rapid changes. An artificial source of illumination is therefore provided and is thrown upon the deposition surface by means of an elliptical reflector. The ellipse, which is cut out of $\frac{1}{4}$ " thick plate glass, is shown in plan in figures 3 and 4. The light is placed in a hole cut at one focus of the ellipse and the thimble fits in a conical seat in a hole at the other. The edges of the ellipse are silvered so that an image of the lamp is formed over the thimble. Loss of light from the top and bottom surfaces of the ellipse is prevented by total internal reflection. The hole into which the thimble fits is made conical in opposing senses above and below the thimble top. The lower cone provides a seating for the thimble top which must be a good fit to prevent water being drawn in by capillary action, and the upper cone provides a prism to refract light down on to the thimble surface to illuminate the deposit. Both cones are cut to an angle of $17\frac{1}{2}^\circ$. The total angle of the two cones must be less than 37° if light reflected by one surface is to be totally internally reflected back into the ellipse by the other. If this is not done the thimble is seen surrounded by a dazzling ring of light.

The hole in the ellipse over the thimble forms a chamber into which the air is supplied by the jet. A slot, $\frac{1}{4}$ " wide, cut in the ellipse is seen in the plan of the ellipse in figures 3 and 4, and the brass jet is brought in through this slot. The brass body of the jet is $\frac{1}{4}$ " high, but it is only 0.2" wide so that there is a gap on either side to provide an outlet for the air which enters the chamber through the jet. This arrangement with the outlet channels on either side of the jet, was shown, by the use of smoke, to give the maximum scavenging from the chamber. If the outlet hole is opposite the jet, vortices are formed on either side of the jet, in which stale air can remain.

The oblique illumination which the ellipse provides gives an effect similar to dark ground illumination and to enhance this effect the anodized surface is dyed black giving a fine black surface as free as possible from sheen; this surface is very hard and can be cleaned without scratching. This arrangement gives the maximum contrast between a single ice crystal and its background, but, by using a small jet which fully ventilates only a narrow streak across the thimble the observed contrast is further increased, since the eye can compare the area of the



Plan view—cover plate removed

Figure 4. Frost-point hygrometer for pressure cabin aircraft, diagrammatic illustration
(Reproduced, by kind permission, from the 'Journal of the Royal Aeronautical Society')

jet in which deposit is at a maximum with the rest of the thimble which usually has little or no deposit. At very low temperatures where the individual ice crystals are too small to be seen, the eye is dependent upon this latter effect to perceive the deposit.

To assist the eye further a moderate amount of optical magnification is provided. It is necessary to see the whole of the thimble top at once in order to permit comparison between the deposit in the jet and the rest of the thimble. The thimble top is 15 mm. diameter and a magnification of more than $\times 10$ would prevent simultaneous vision of the whole thimble. Actually a magnification of $\times 7$ is provided. It is not possible to use an ordinary magnifying glass or eyepiece for this purpose partly because of the absence of any particular eye point and partly because the bottom surface of the lower lens must form the top of the observation chamber, and it must therefore be as low as possible. It must be, however, lifted sufficiently for it not to be illuminated by the ellipse. If it is illuminated any dust or dirt on the bottom surface is very visible. The optical system adopted, figure 4, is therefore corrected after the manner of a Petzval portrait lens rather than a conventional eyepiece. The upper lens, an achromat of B.S.C. and L.D.F. glasses, is corrected for spherical aberration, it has slight positive coma and an inevitably large amount of positive astigmatism which would be very undesirable if uncorrected. This particular combination of glasses is used to permit achromatization with relatively long radii and therefore to permit a large diameter on the lens. This large size of the upper lens permits the eye to be moved at least 2" from the upper lens and still retain a full view of the thimble. The lower lens is made an approximate achromat of H.C. and D.F. glasses, largely because the use of a doublet in this position greatly assists in the correction of the astigmatism of the upper lens. The combined system has no visible spherical aberration or coma. It has an almost flat tangential field for any reasonable position of the eye, and no visible chromatic aberration.

The combination of dark ground illumination and good quality optical magnification shows a crystal of ice such as is obtained at temperatures down to -60°C . in brilliant contrast. At very low temperatures when the individual crystals cannot usually be seen the system shows up a deposit consisting of about 0.01 microgram of water as a bluish bloom in the line of the jet.

Misting of the lenses is apt to be troublesome since very cold air is circulating in the chamber under the lower lens while the cabin temperatures may be quite high. This is particularly troublesome in the pressurized instrument which is commonly used in a pressure cabin, probably with a cabin temperature of 15 or 20°C . and a high relative humidity, maintained by perspiration and exhalation by the aircraft crew, while the air under the lower lens may be -80°C . (In the stratosphere it is possible with the same instrument to measure the CO_2 point, in which case the air in the test-chamber may be still colder.) This trouble is eliminated by making the tube which carries the upper lens of plastic so that thermal conduction is reduced and the upper lens is not cooled sufficiently for condensation to take place upon it. The lower lens mount, on the other hand, is deliberately made of brass or other good thermal conductor so that the low temperature of the test-chamber is readily conducted through the mount to the space between the lenses. Any water vapour in the space between the lenses then condenses upon the cold mount rather than upon the glass surface, which is slightly warmer due to the reduced conduction in the glass. To reduce con-

duction through the glass, the lower lens is made very thick. At a total thickness of the lower lens of 0.25" some trouble was experienced in a pressure cabin, but by increase of thickness to 0.5" condensation on the lenses was entirely eliminated.

Absorption of water by various materials is always a source of error as it is given up again when working with very dry air. Cleaned copper tubing is the most satisfactory material for supply pipe lines and flexible metal bellows must be used to take up any movement; rubber tubing must always be avoided. For the same reason it has been necessary to make the illuminating ellipses of glass, though the difficulties of manufacture are increased. Perspex is very suitable from the point of view of manufacture but it is useless if frost-points below -30°C . are to be measured. Transpex II (polystyrene) is less satisfactory for manufacture, but can be used down to about -60°C . The glass ellipse is satisfactory for use at all temperatures.

To use the instrument and obtain an accurate reading at low temperatures in a reasonable time considerable experience is required. The usual method is to pump slowly so that the thimble temperature falls at a reasonable rate, usually less than $\frac{1}{2}^{\circ}\text{C}$. sec., watching carefully for the formation of a deposit. The deposit does not necessarily begin to form immediately the frost-point is passed, so that the deposit may not be seen until after a quite appreciable overshoot. Immediately a deposit is seen, pumping is stopped and the thimble allowed to warm up one or two degrees, depending upon the rate of cooling which has been used, upon the actual frost-point, and upon the particular thimble, which may be very clean or may have been sensitized by past use. The thimble temperature is then held constant by careful pumping and the deposit examined intermittently to see if it is increasing or decreasing. With experience it is then easy to find two temperatures at which the deposit is (a) very slowly increasing, and (b) very slowly decreasing, at approximately the same rates, and the dew- or frost-point is taken as the mean. On the other hand, if care is not taken the deposit may be lost altogether or may become too heavy. In either case much time is lost in obtaining a suitable deposit again. In order to get accurate measurements the deposit should be very thin since if it becomes at all heavy, changes are difficult to observe. It is very important that a deposit should be maintained throughout the observation and changes in the deposit should be observed while the thimble temperature is held steady. The frost-point is *not* the mean of the temperatures at which the deposit comes and goes from a thimble whose temperature is fluctuating rapidly, and this tempting method of use must be strictly avoided.

Laboratory tests of the accuracy of the instrument. At frost-points down to -50°C . the accuracy which is obtainable is largely determined by the accuracy with which the temperature of the thimble can be controlled, and it is not difficult to measure a frost-point of -50°C . repeatedly to $-\frac{1}{2}^{\circ}\text{C}$. in the laboratory, since the temperatures of deposit increase and decrease may differ by less than, say, 2°C . This accuracy can be obtained in an aircraft if care can be taken, though usually it is convenient to accept a less accurate observation which can be made more quickly and with much less effort. The humidity of the air through which an aeroplane is flying sometimes varies quickly so that great accuracy is not necessary.

Tests were made of the fundamental accuracy of the hygrometer at -78.5°C .

to determine (a) whether the frost-point as measured was the true saturation temperature, (b) whether there was any effect on the measured frost-point when the thimble surface was changed from gold to the black anodised aluminium surface, and (c) the probable error of an observation at this temperature, and whether this was different for the two kinds of deposition surface or whether there was any major difference between various experienced observers.

Air with a frost-point of -78.5°C . was produced by first passing it through large drying tubes of silica gel at a pressure of 120 lb./sq.in. and expanding to atmospheric pressure; this gives air with a frost-point of about -65°C . The air was then passed through two spirals of clean copper tube immersed in a bath of solid carbon dioxide and acetone, which, provided it is boiling at a reasonable rate, gives a satisfactory constant temperature of -78.5°C . The air from the spirals was next passed through a large glass-wool filter also immersed in a carbon dioxide and acetone bath; the filter is used to remove particles of snow which are carried along by the air-flow and its use is essential. Provided that the system is not expected to remove too much water and that there is no excessive pressure difference across the filter, this arrangement gives air of a very constant water content and its frost-point should be -78.5°C ., the temperature of the baths. The air was finally passed to a hygrometer where its frost-point was measured. In the laboratory a copper-constantin thermocouple inserted into the thimble under the test surface was used to measure the thimble temperature, and for this experiment the other junction was placed in the cooling bath surrounding the filter so that the measured frost-point should be at zero E.M.F.

The results obtained in one set of readings are shown in table 2.

Table 2

Observer	Temperature difference, Thimble minus Bath		Mean
	Deposit just growing	Deposit just decreasing (Galvanometer deflection)	
Black aluminium thimble			
A.W.B.	-2.5 div.	+ 3.5 div.	+ 0.5 div.
B.C.	1.5	4.5	1.5
G.M.B.D.	-0.5	+ 1.5	+ 0.5
			— — —
			Mean 0.8 div.
Gold topped thimble			
A.W.B.	2.5 div.	+ 1.5 div.	0.5 div.
B.C.	1.5	1.5	0
G.M.B.D.	-1.5	1.5	0
			— — —
			Mean 0.2 div.
Black aluminium thimble			
G.M.B.D.	0.5 div.	+ 3.5 div.	+ 1.5 div.
B.C.	1.5	+ 1.5	0
			— — —
			Mean + 0.7 div.

(Note. Thimble temperatures were held constant for about a minute at various whole divisions on the galvanometer, and at the end of the minute the observer decided whether the deposit had grown or decreased. The galvanometer had a zero error of 0.5 divisions, therefore all values are given to this accuracy. 1 div. $\sim 1^{\circ}\text{C}$. Each complete observation took about five minutes.)

It will be seen that the frost-points as observed by the three observers agree to within about 1°C . and that there is no significant difference between a gold and a black anodized aluminium thimble, nor is there significant difference between the expected and measured frost-point. In general most observers have a slight preference for the black anodized thimble, and this is used for the final instrument mainly because the provision of a resistance thermometer is so much easier.

Air with a frost-point as low as -78.5°C . is usually only encountered at heights in the atmosphere where, owing to shortage of oxygen and to discomfort, it is impossible to make the observations with the care possible in the laboratory; the more accurate type of instrument described below should therefore be used at very low temperatures.

§ 4. FROST-POINT HYGROMETER WITH PHOTOELECTRIC DEPOSIT INDICATOR

The eye-observation frost-point hygrometer described in the previous section makes a very simple instrument and is easy to use when measuring air whose frost-point is above about -50°C . When it has to be used to measure the humidity of still drier air it suffers from serious disadvantages. Although it is possible to measure the frost-point of very dry air down to -80°C ., the rate of deposition and evaporation of the hoar frost become very slow so that it is necessary to keep the thimble at a constant temperature for about a minute before it is possible to decide whether the deposit is increasing or decreasing. At -80°C . the deposit is very faint indeed. This makes the measurements very long and fatiguing to the observer and the accuracy is rather poor. An instrument with a photoelectric deposit indicator has been developed to remove these disadvantages, and though it is necessarily more complicated than the simple eye-observation instrument, it is so much more accurate, so much quicker, and so much less tiring to use, that its use is always worth while at very great heights. Also, since much less judgment is required, the observations may be made by a less highly trained observer. Photoelectric dew-point hygrometers have been described by other workers, including Moss (1934), Awbery and Griffiths (1933), Thornthwaite and Owen (1940), and others in the U.S. and Canada. The optical arrangements employed are generally unsuitable for the lowest frost points.

The principle of the instrument is seen from the schematic diagram

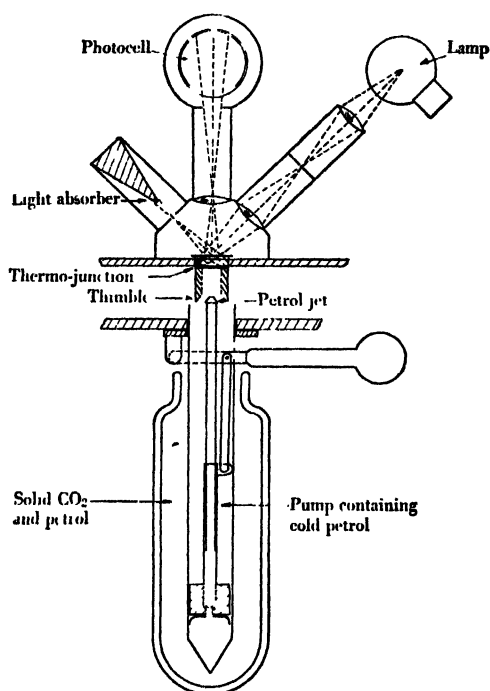


Figure 5. Diagram showing the principle of the frost-point hygrometer with photoelectric deposit indicator
(Reproduced, by kind permission, from the 'Proceedings of the Royal Society'.)

in figure 5. Light from a 36-watt lamp is concentrated by lenses on to the thimble. The thimble is made of copper with a gold plate soldered to its top, the gold surface being well polished or of aluminium with an anodized black top. The light is reflected from the thimble into a black absorbing box, through a slot slightly larger than the image of the lamp filament. A slot between the two lenses is also larger than the image of the filament and prevents the passage of all light except that directly from the filament. A diaphragm of suitable aperture cuts out reflections from the inside of the tube in which the lenses are mounted. As a result the inside of the chamber is fairly dark, except that an area about 8 mm. in diameter on the top of the thimble is strongly illuminated. A lens immediately above the thimble forms an image of this illuminated area at a hole through which the light passes to the photoelectric cell. The lenses are all ventilated by a stream of air from within the test-chamber to prevent any condensation on them, as they will tend to be warmer than the outside air.

If the surface of the thimble is clean and properly polished very little light is scattered from it on to the photocell. It is important both in the final polishing and in subsequent cleaning that all rubbing of the surface be only in the plane of the light since any minute scratches at right angles to this direction show up brightly. A small deposit of dew or hoar frost on the thimble causes a great

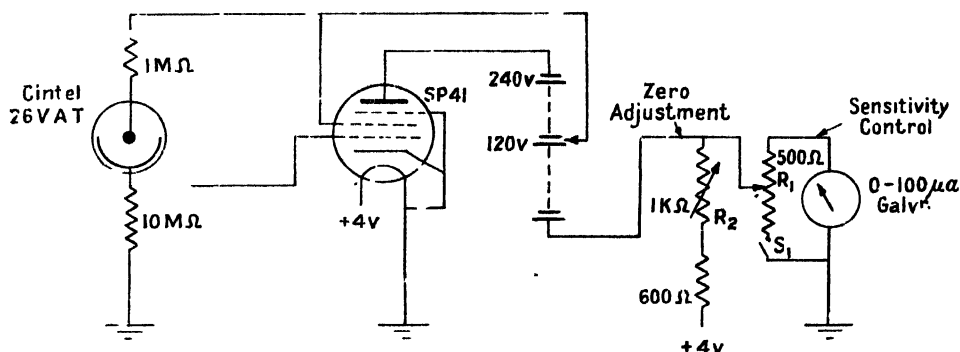


Figure 6.

relative increase in the amount of light reaching the photocell. The angle of the illuminating beam and the position of the photocell were fixed after careful trials using very faint deposits formed in very dry air. Somewhat different positions might be better for heavy deposits which appear white, but there is always ample light in these cases and the positions were chosen which were best for the very faint bluish deposits (see § 2).

The electric circuit is shown in figure 6. With a grid leak of resistance greater than 10^7 ohms the instrument is too sensitive, even when measuring air having a frost-point of -80°C . The instrument is made to be of maximum sensitivity when used with frost-points of -80°C . and when measuring higher frost-points the sensitivity is decreased by adjusting the resistance R_1 . The indicating microammeter is never shorted by a resistance less than several times its own resistance so that it shall not be too sluggish. For full sensitivity the switch S_1 is opened. By adjusting the resistance R_2 the current through the indicating microammeter is made approximately zero when there is no deposit on the thimble. When using very dry air it is found that the change in the amount of

scattered light produced by a given small increase in the deposit is greatest for a moderately thin, but not very thin, deposit. With the photocell and amplification used, a deposit giving a reading of a few hundred microamperes is most sensitive. It is convenient to have an indicating galvanometer with high sensitivity in the middle of its range (when the current is about a hundred microamperes) and becoming less sensitive at each end of the scale.

The air to be measured is led in through a slot 10 mm. \times 1 mm., the centre of the jet being directed at the centre of the thimble and making an angle of approximately 10° with the surface, as in the visual hygrometer. The air jet more than covers the illuminated area and is directed at right angles to the plane of the illumination.

The thimble may be cooled by pumping a jet of liquid air against the underside, as in the case of the visual instrument described above; or it may be cooled by being connected by a short, thin bar to a massive copper bar passing down into a Dewar flask of liquid air. In the latter case a heating coil is wound on the side of the thimble in a manner similar to the winding for the resistance thermometer already described. The current through the heating coil is controlled in order to keep the thimble at the required temperature. The photoelectric instrument is chiefly used when climbs are made to very great heights, in which case aircraft with pressurized cabins are generally used. The main instrument is then put outside the pressure-cabin and controlled from inside. If the jet method of cooling is used, the pump must be worked by means of a Bowden cable or other similar method, and the conductivity method of cooling with controlled heating may be more convenient.

Method of use.—The method of using the photoelectric instrument is different from that for the visual instrument. Since it is very difficult to judge by eye whether a deposit is keeping exactly constant, it was found necessary to bring the thimble to two temperatures where the deposit was definitely slowly increasing and slowly decreasing, when the true frost-point would be near the mean. With the photoelectric instrument the indications of the deposit are so sensitive that it is possible to say when the deposit is neither increasing nor decreasing, and the temperature of the thimble is adjusted to such a value and the temperature read at this condition. As the actual amount of the deposit is not important (provided it is not too heavy), many readings can be taken quickly one after the other with slightly different amounts of deposit. At least two independent measurements can be taken per minute so that a very reliable average can quickly be obtained.

Since the thimble cannot be seen by the observer there may be some doubt whether the deposit consists of dew or frost at temperatures between 0° and about -30°C . Since dew consists of very numerous fine droplets, while frost at these temperatures consists of much larger ice particles, a deposit of dew increases and decreases much more rapidly than one consisting of frost, and a skilled observer may judge the nature of the deposit from this. However, it is generally safer to cool the thimble sufficiently to be sure that the deposit has frozen, then to warm it until nearly all the ice has evaporated, leaving only a very thin deposit, and to work with this.

Measurements of frost-point can be made with this instrument down to the temperature at which the nature of the deposit changes from a crystalline to an invisible glassy layer.

The following critical tests of the instrument have been made in the laboratory with satisfactory results: —

A. To see what effect the thickness of the deposit had on the observed frost-point.

Observations were made using a very thin, a moderate and a very thick deposit, the order of the observations being shown by the figures within brackets (table 3).

Table 3

Observed frost-points (°c.)					
Thin deposit		Moderate deposit		Thick deposit	
			— 57.2		— 58.6
		(1)	— 57.1	(2)	— 57.8 ?
			57.3		— 58.6
		Mean	57.2	Mean	— 58.4
	57.5		— 58.2		— 57.8
(3)	57.5	(4)	58.0	(5)	— 58.0
	— 57.3		58.1		— 58.2
Mean	— 57.5	Mean	— 58.1	Mean	— 58.0
	58.0		— 58.3		58.5
(6)	— 58.1	(7)	58.4	(8)	— 58.3
	— 58.3		58.7		58.7
Mean	— 58.1	Mean	— 58.5	Mean	— 58.5
	— 58.2				
(9)	— 58.2				
	58.2				
Mean	— 58.2				
Means	— 57.9		— 57.9		— 58.3

The air supply was obtained by passing air over silica gel. It will be noticed that the air supply becomes slowly drier as the connecting pipes dry out. All the individual measurements are given so that the casual variations can be seen.

Since the thick and thin deposits used on the above test were outside the range of thickness usually used, no appreciable change of observed frost-point with deposit thickness is likely.

B. To see if the brightness of the illuminating lamp caused any change in the observed frost-point (e.g. through heating of the deposit on the thimble).

For this test the lamp was run alternately on 4 and 6 volts (the normal rating), while many measurements of the frost-point were made.

The means of the observed values were -50.7° c. for the 6-volt run and 50.2° c. for the 4-volt. Since the energy radiated by the lamp is much greater when on 6 volts than on 4 volts (which was the minimum on which the instrument would work satisfactorily), it may be taken that there is no appreciable effect.

C. To see if the rate of air-flow through the jet makes any difference to the observed frost-point.

The pressure on the jet was adjusted alternately to 4 in. of water and 24 in. with the following results (table 4).

Table 4

Observed frost-points (°C.)

Pressure on jet			
24 in.		4 in.	
	— 54·8		— 55·3
(1)	— 54·5	(2)	— 55·2
	— 54·8		— 55·2
	— — —		— — —
Mean	— 54·7	Mean	— 55·2
	— 56·2		— 56·2
(3)	— 56·0	(4)	— 56·0
	— 56·0		— 55·9
	— — —		— — —
Mean	— 56·1	Mean	— 56·0
	— 56·2		
(5)	— 56·0		—
	— 56·0		
	— — —		
Mean	— 56·1		
Means	— 55·6		— 55·6

Note that the air supply is becoming drier during the first part of the test, probably due to the tubing drying out.

ACKNOWLEDGMENTS

This work was done while two of us (A.W.B. and B.C.) were on the staff of the Meteorological Office, and we are indebted to the Director both for permission to publish this paper and for his personal interest in the progress of the work.

Figures 1, 2, 3 and 5 are reproduced by permission of the Council from the *Proceedings of the Royal Society*, and figure 4 from the *Journal of the Royal Aeronautical Society*, by permission of the Council of that Society.

REFERENCES

- AWBERY and GRIFFITHS, 1933, *The Measurement of Humidity in closed spaces*. D.S.I.R., Food Investigation Report No. 8 (London : H.M.S.O.).
- DIAMOND, HINMAN, DUNMORE and LAPHAM, 1940, *Bur. Stand. J. Res., Wash.*, **25**, 327.
- DYMOND, 1947, *Proc. Phys. Soc.*, **59**, 645.
- EBERT, 1937, *Z. f. Kälte-Ind.*, **44**, 127.
- GLÜCKAUF, 1945, *Quart. J. R. Met. Soc.*, **71**, 110.
- GREGORY, 1947, *Instrum. Practice*, **1**, 367.
- GRIFFITHS and AWBERY, 1935, *Proc. Phys. Soc.*, **47**, 684.
- HARRISON and BREWER, 1944, *Meteorological Research Committee Reports*, M.R.P. 205.
- MOSS, 1934, *Proc. Phys. Soc.*, **46**, 450.
- SIMONS, 1936, *Proc. Phys. Soc.*, **48**, 136.
- THORNTHWAITE and OWEN, 1940, *Mon. Weath. Rev., Wash.*, **68**, 315.

The Cavity Resonator Method of Measuring the Dielectric Constants of Polar Liquids in the Centimetre Band

BY C. H. COLLIE, J. B. HASTED AND D. M. RITSON

Clarendon Laboratory, Oxford

M.S. received 24 January 1947

ABSTRACT. A method is described of determining the refractive index n of a polar liquid in the centimetric band by measuring the resonance curve of an H_{01} resonator with an axial capillary of liquid. An independent value of κ is required. The method has been tried out on water at 21°C. Three independent measurements gave $n = 6.25, 6.30, 6.35$, so that the method is capable of $\pm 1\%$ accuracy.

§ 1. INTRODUCTION

THE most convenient way of measuring the dielectric properties of nearly ideal dielectrics (power factor 10^{-3}) is the cavity resonator method introduced by Willis Jackson (1941), using an E_{010} resonator, and Penrose (1946), using an H_{01} resonator. The latter method is in some respects more convenient at the highest frequencies ($\lambda = 1$ cm.), since it is possible to work with an oscillator of fixed frequency.

It is our purpose to examine whether these methods can be used when the dielectric is a polar liquid such as water, which is far from being a perfect dielectric (power factor $\frac{1}{2}$), and furthermore, which has to be maintained in the appropriate form by a containing vessel. If the problem is considered by an approximate method due to Bethe (1942), it is easily seen that it is not possible to arrange the specimen so as to obtain a large frequency shift and still maintain a sharp enough resonance curve for the frequency shift to be measured.

Two main difficulties are encountered in a straightforward application of the resonator method. In the case of the E_{010} resonator, in which the electric field is greatest on the axis, the loss is such that at 1 cm. the specimen must be enclosed in a capillary of less than 0.05 mm. diameter, which is impracticable. Moreover, it is difficult to achieve 1% accuracy with a fixed resonator and tunable oscillator. In the case of the H_{01} resonator, the difficulty is one of introducing the liquid specimen in the form of a flat disc of thickness 0.2 mm. These difficulties can to a great extent be overcome by using an H_{01} resonator with an axial capillary filled with liquid. The electric field is zero on the axis of the resonator so that a capillary tube large enough for accurate measurement can be used without damping the resonator so much that the shape of the resonance curve cannot be determined. The cylindrical symmetry makes it possible to solve Maxwell's equations exactly, so that the small corrections, which have to be made to take into account the effect of the glass wall of the capillary tube, can be evaluated as accurately as one wishes.

§ 2. THE RESONANCE CURVE AND THE PROPAGATION CONSTANT

The shape of the response curve of a resonator can be obtained by considering the course of a wave travelling back and forth in the resonator (e.g. Collie (1944)).

Consider the H_{01} resonator of length $l/2$ shown in figure 1.

If precautions are taken that new types of wave are not created and subsequently detected, the effect of the plunger and the blank face and detector at the input end can be represented by complex reflexion coefficients R_p and R_t respectively.

Then the electric field at the detector due to waves travelling from left to right is

$$A(1 + R_p R_t e^{-\gamma l} + (R_p R_t)^2 e^{-2\gamma l} + \dots) = \frac{A}{1 - R_p R_t e^{-\gamma l}},$$

while the field at the detector of waves travelling from right to left is

$$B e^{-\gamma l} R_p (1 + R_p R_t e^{-\gamma l} + \dots),$$

so that the total electric field E at the detector is given by

$$E = \frac{A + B R_p e^{-\gamma l}}{1 - R_p R_t e^{-\gamma l}},$$

in which γ is the propagation constant $\alpha + j\beta$ of the wave in the guide.

Near resonance let

$$l = l_0 + \Delta l,$$

$$\beta l_0 = 2n\pi - \phi,$$

$$R_p R_t = e^{-r - j\phi},$$

$$B R_p e^{-\gamma l_0} = C.$$

Then

$$\begin{aligned} E &= \frac{A + C e^{-\gamma \Delta l}}{1 - e^{-r - j\phi - \gamma l}} \\ &= \frac{A + C e^{-\gamma \Delta l}}{(r + \alpha l) - j\beta \Delta l}, \end{aligned}$$

in which A can be treated as real without affecting the argument.

Then $|E|$ is given by

$$|E| = \frac{F + G \Delta l}{\{(r + \alpha l)^2 + \beta^2 \Delta l^2\}^{1/2}}.$$

This is a conventional resonance curve made slightly asymmetric by the small linear term $G \Delta l$.

The error involved in treating it as a conventional Q curve and measuring $\frac{1}{2}$ (full width) when $E = E_{\max}/\sqrt{2}$ is quite small; if the degree of asymmetry is 5% at the half value width, the error in estimating $(r + \alpha l)$ is only 0.15%. A more satisfactory graphical treatment is to plot the observed value of Δl^2 against the square of the reciprocal of the mean value of the voltage response. The constant $r + \alpha l$ is readily obtained from the slope and intercept of the resulting straight line.

A typical response curve of a damped resonator is shown in figure 2 together with the linear plot described above. It can be seen that an excellent straight line is obtained: this not only justifies the elementary theory just outlined but provides a good check on the linearity of the receiving system.

The quantities required in theoretical calculations are α and β . These can be obtained if two successive resonances are measured. The amount the plunger has to be withdrawn is a direct measure of the wavelength λ_g in the guide and so of $\beta = 2\pi/\lambda_g$. Since r is the same for both positions of the plunger, α can be obtained by subtraction from the measured values of $r + \alpha l_0$ and $r + \alpha(l_0 + \lambda_g/2)$. Thus, provided the resonance curve is not too flat for accurate measurement, one can determine γ , the propagation constant in the guide, experimentally from resonance data.

3. PROPAGATION IN A COMPOSITE GUIDE

It remains to obtain the relationship between the propagation constant γ in a guide partly filled with dielectric and the constants of the material. A general treatment of this problem is given by Pincherle (1944). Consider the propagation of an H_{01} wave down to the composite guide shown in figure 3, consisting of a meta

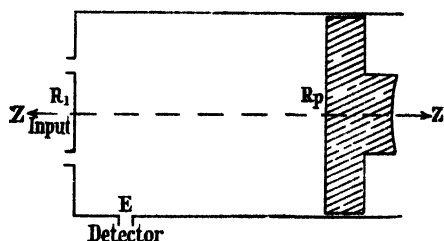


Figure 1.

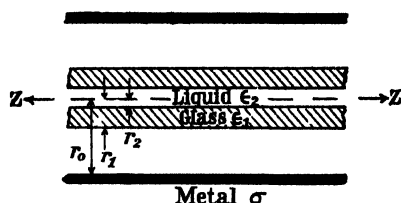


Figure 3

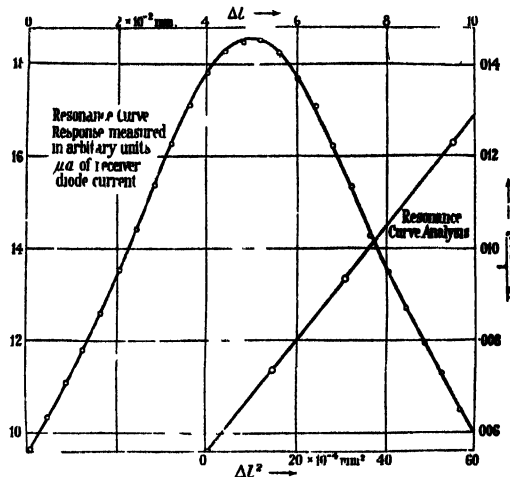


Figure 2 Resonance curve plotted and analysed as a straight line.

cylinder of internal radius r_0 , with a glass capillary tube of external radius r_1 , and internal radius r_2 , lying along its axis. Let σ be the conductivity of the metal wall and ϵ_1 and ϵ_2 the complex dielectric constant of the glass and liquid respectively. Then the field equations for an H_{01} wave in cylindrical coordinates ρ, θ, z are *

$$H_z = k^2 [a J_0(k\rho) + b Y_0(k\rho)], \quad \dots\dots (1)$$

$$H_\rho = -\gamma k [a J_1(k\rho) + b Y_1(k\rho)], \quad \dots\dots (2)$$

$$E_\theta = \frac{j\omega\mu k}{c} [a J_1(k\rho) + b Y_1(k\rho)], \quad \dots\dots (3)$$

$$k^2 - \gamma^2 = \frac{\omega\epsilon\mu}{c^2}. \quad \dots\dots (4)$$

The arbitrary complex coefficients a and b will have different values (cd, ef, gh) in each medium which must be chosen so that the tangential electric and magnetic fields are continuous at each interface.

* Lamont, *Wave Guides* (Methuen), p. 26; his $\beta = -j\gamma$.

One notes that γ must be the same for all four sections of the wave since these boundary conditions must continue to hold as the wave is propagated down the guide. Since γ is a measured quantity, the appropriate value of k (k_M, k_A, k_U, k_L) can be determined from equation (4) except in the inner region which is filled with the liquid whose dielectric constant ϵ_2 is being measured.

Approximations

The following well known approximations will be used:

For $a \ll 0.01$

$$A(1) \quad J_1(3.831 + a) \approx a J'_1(3.831),$$

$$A(2) \quad J_0(3.831 + a) \approx J_0(3.831),$$

$$A(3) \quad Y_0(3.831 + a) \approx Y_0(3.831),$$

$$A(4) \quad Y_1(3.831 + a) \approx Y_1(3.831).$$

For $x \leq 3.8$ and $a/x \ll 0.01$

$$A(5) \quad J_1(x + a) \approx J_1(x),$$

$$A(6) \quad J_0(x + a) \approx J_0(x),$$

$$A(7) \quad Y_1(x + a) \approx Y_1(x),$$

$$A(8) \quad Y_0(x + a) \approx Y_0(x).$$

For $b \ll 0.01$

$$A(9) \quad \frac{J_1(x) + b Y_1(x)}{J_0(x) + b Y_0(x)} \approx \frac{J_1(x)}{J_0(x)} + b \frac{Y_1(x) \cdot J_0(x) - Y_0(x) \cdot J_1(x)}{[J_0(x)]^2}.$$

In the metal

Since the fields must be zero for large values of ρ , $b = -ja$ and the field in the metal can be written in terms of Hankel functions as

$$H_z = a k_M^2 H_0^{(2)}(k_M \rho) e^{-\gamma \rho}, \quad \dots (5)$$

$$H_\rho = -a \gamma k_M H_1^{(2)}(k_M \rho) e^{-\gamma \rho}, \quad \dots (6)$$

$$E_\theta = \frac{a j \omega k_M \mu}{c} H_1^{(2)}(k_M \rho) e^{-\gamma \rho}, \quad \dots (7)$$

in which k_M denotes k in the metal given by

$$k_M^2 = \frac{4\pi\sigma\omega}{c^2} + \gamma^2 \approx \frac{4\pi\sigma\omega}{c^2}. \quad \dots (8)$$

Since the value of γ is never very different from that of an empty guide with perfectly conducting walls it is convenient to write our expressions in terms of their deviation from the propagation constant G of an ideal empty guide, given by

$$G^2 = \frac{3.831^2}{r_0^2} - \omega^2/c^2.$$

The wave in the actual empty guide will be given by

$$H_z = a_0 \cdot J_0(k_A \rho) \cdot J_0(k_A r_0), \quad \dots (9)$$

in which the subscript zero indicates values for an empty guide.

From (8) it follows that $k_M \approx k_M$, so that equating E_θ and H_z at the metal/air interface of an empty guide, one has

$$\frac{1}{k_M r_0} \frac{H_1^{(2)}(k_M r_0)}{H_0^{(2)}(k_M r_0)} = \frac{1}{k_A r_0} \cdot \frac{J_1(k_A r_0)}{J_0(k_A r_0)}. \quad \dots (10)$$

$$\text{Let } k_A r_0 = 3.831 + \chi_1 + j\chi_2, \quad \dots\dots(11)$$

in which the χ s represent the effect of the finite conductivity of the wall and are very small.

$$\text{Let } \gamma_0 = G + \gamma_{01} + j\gamma_{02}, \quad \dots\dots(12)$$

so that from (4), (11), (12) one has

$$\chi_1 + j\chi_2 \simeq \frac{r_0^2}{3.831} G(\gamma_{01} + j\gamma_{02}), \quad \dots\dots(12a)$$

and substituting in (10)

$$\frac{1}{k_M r_0} \cdot \frac{H_1^{(2)}(k_M r_0)}{H_0^{(2)}(k_M r_0)} \simeq \frac{r_0^2}{(3.831)^2} \cdot G \cdot \frac{J_1'(3.831)}{J_0'(3.831)} (\gamma_{01} + j\gamma_{02}). \quad \dots(13)$$

The air metal boundary

The boundary conditions are expressed by the equations

$$ak_M^2 H_0^{(2)}(k_M r_0) = k_A^2 [cJ_0(k_A r_0)] + dY_0(k_A r_0), \quad \dots\dots(14)$$

$$ak_M H_1^{(2)}(k_M r_0) = k_A [cJ_1(k_A r_0)] + dY_1(k_A r_0). \quad \dots\dots(15)$$

These can be solved for d/c by division; $k_A r_0$ is nearly 3.831, so writing

$$k_A r_0 = 3.831 + \phi_1 + j\phi_2$$

and using approximations A (1), A (2), A (3), A (4) and neglecting $(d/c) \cdot Y_0(3.831)$ compared with $J_0(3.831)$, one has

$$\frac{1}{k_M r_0} \left(\frac{H_1^{(2)}(k_M r_0)}{H_0^{(2)}(k_M r_0)} \right) \simeq \frac{1}{3.831} \left[(\phi_1 + j\phi_2) \cdot \frac{J_1'(3.831)}{J_0'(3.831)} + \frac{d}{c} \cdot \frac{Y_1(3.831)}{J_0(3.831)} \right],$$

substituting for the L.H.S. from equation (13) and writing as before

$$\gamma = G + \gamma_1 + j\gamma_2,$$

$$\text{so that } G(\gamma_1 + j\gamma_2) \simeq \frac{3.831}{r_0^2} (\phi_1 + j\phi_2)$$

as in (12a), one finds that

$$\frac{d}{c} \simeq - \frac{J_1'(3.831)}{Y_1(3.831)} \cdot \frac{r_0^2 G}{3.831} \cdot (\gamma - \gamma_0) = u(\gamma - \gamma_0)$$

$$\text{with } u = \frac{-J_1'(3.831)}{Y_1(3.831)} \cdot \frac{r_0^2 G}{3.831}, \quad \dots\dots(16)$$

u has no real part since G is purely imaginary.

The glass air surface

The boundary conditions are

$$k_A^2 [cJ_0(k_A r_1) + dY_0(k_A r_1)] = k_G^2 [eJ_0(k_G r_1) + fY_0(k_G r_1)], \quad \dots\dots(17)$$

$$k_A [cJ_1(k_A r_1) + dY_1(k_A r_1)] = k_G [eJ_1(k_G r_1) + fY_1(k_G r_1)], \quad \dots\dots(18)$$

$$\text{so that } \frac{1}{k_A} \cdot \frac{J_1(k_A r_1) + (d/c)Y_1(k_A r_1)}{J_0(k_A r_1) + (d/c)Y_0(k_A r_1)} = \frac{1}{k_G} \frac{J_1(k_G r_1) + (f/e)Y_1(k_G r_1)}{J_0(k_G r_1) + (f/e)Y_0(k_G r_1)}, \quad \dots\dots(19)$$

which can be solved for f/e since d/c can now be looked upon as known.

Making use of the relationship $J_1(x)Y_0(x) - J_0(x)Y_1(x) = 2/\pi x$ one finds that

$$\frac{k_{\text{L}}}{k_{\text{A}}} \left[\frac{J_1(k_{\text{A}}r_1)}{J_0(k_{\text{A}}r_1)} - \frac{d}{c} \cdot \frac{2}{J_0^2(k_{\text{A}}r_1) \cdot \pi k_{\text{A}}r_1} \right] \simeq \frac{J_1(k_{\text{A}}r_1)}{J_0(k_{\text{A}}r_1)} - \frac{f}{e} \frac{2}{J_0^2(k_{\text{G}}r_1) \cdot \pi k_{\text{G}}r_1} \quad \dots\dots (20)$$

This may be further reduced using A (5, 6, 7, 8), which are valid if the glass is nearly loss-free, so that $\epsilon'_{\text{G}} \sim \epsilon''_{\text{G}}/100$.

For the Corning 707 low-loss glass used, this approximation is justified.

One finds
$$\frac{f}{e} = v + w d/c \quad \dots\dots (21)$$

with
$$v = -\pi \bar{k}_{\text{G}} r_1 \left[\frac{\bar{k}_{\text{A}}}{\bar{k}_{\text{G}}} \cdot \frac{J_1(\bar{k}_{\text{A}}r_1)}{J_0(\bar{k}_{\text{A}}r_1)} - \frac{J_1(\bar{k}_{\text{G}}r_1)}{J_0(\bar{k}_{\text{G}}r_1)} \right] J_0(\bar{k}_{\text{G}}r_1), \quad \dots\dots (22)$$

$$w = \frac{\bar{k}_{\text{G}}^2}{k_{\text{A}}^2} \cdot \frac{J_0^2(\bar{k}_{\text{A}}r_1)}{J_0^2(\bar{k}_{\text{A}}r_1)} \quad \dots\dots (23)$$

in which \bar{k}_{G} and \bar{k}_{A} denote the real parts of k_{G} and k_{A} respectively.

The glass liquid surface

The boundary conditions are

$$k_{\text{G}}^2 [e J_0(k_{\text{G}}r_2) + f Y_0(k_{\text{G}}r_2)] = k_{\text{L}}^2 g J_0(k_{\text{L}}r_2), \quad \dots\dots (24)$$

$$k_{\text{G}} [e J_1(k_{\text{G}}r_2) + f Y_1(k_{\text{G}}r_2)] = k_{\text{L}} g J_1(k_{\text{L}}r_2), \quad \dots\dots (25)$$

there being no term $h Y_0(k_{\text{L}}r)$ since the fields must be finite on the axis.

Using A (9) and neglecting the glass as before, one has

$$\frac{1}{k_{\text{L}}r_2} \cdot \frac{J_1(k_{\text{L}}r_2)}{J_0(k_{\text{L}}r_2)} \simeq \frac{1}{\bar{k}_{\text{G}}r_2} \left\{ \frac{J_1(\bar{k}_{\text{G}}r_2)}{J_0(\bar{k}_{\text{G}}r_2)} + \frac{\frac{f}{e} [Y_1(\bar{k}_{\text{G}}r_2)J_0(\bar{k}_{\text{G}}r_2) - Y_0(\bar{k}_{\text{G}}r_2)J_1(\bar{k}_{\text{G}}r_2)]}{J_0^2(\bar{k}_{\text{G}}r_2)} \right\} \quad \dots\dots (26)$$

$$\begin{aligned} \frac{1}{k_{\text{L}}r_2} \cdot \frac{J_1(k_{\text{L}}r_2)}{J_0(k_{\text{L}}r_2)} &= \frac{1}{\bar{k}_{\text{G}}r_2} \cdot \frac{J_1(\bar{k}_{\text{G}}r_2)}{J_0(\bar{k}_{\text{G}}r_2)} + \frac{f}{e} \cdot \frac{2}{\pi(\bar{k}_{\text{G}}r_2)^2} \cdot \frac{1}{J_0^2(\bar{k}_{\text{G}}r_2)}, \quad \dots\dots (27) \\ &= y + \frac{f}{e} z, \end{aligned}$$

so that
$$\frac{1}{k_{\text{L}}r_2} \cdot \frac{J_1(k_{\text{L}}r_2)}{J_0(k_{\text{L}}r_2)} \simeq y + zv + zwu(\gamma - \gamma_0). \quad \dots\dots (28)$$

Determination of the refractive index n

From equation (4) we have

$$k_{\text{L}}^2 - \gamma^2 = \frac{\epsilon_{\text{L}} \omega^2}{c^2} = (n - j\kappa)^2 \frac{\omega^2}{c^2}.$$

For polar liquids ($\epsilon_{\text{L}} \gg 1$) it is sufficiently accurate to write

$$k_{\text{L}} \simeq (n - j\kappa) \omega/c.$$

Substituting into equation (28) and equating real and imaginary parts

$$\Im \left[\frac{c}{\omega r_2(n-j\kappa)} \cdot \frac{J_1(\omega r_2(n-j\kappa)/c)}{J_0(\omega r_2(n-j\kappa)/c)} \right] = zwu \cdot \Re(\gamma - \gamma_0), \quad (29)$$

$$= zwu\alpha,$$

$$\Re \left[\frac{c}{\omega r_2(n-j\kappa)} \cdot \frac{J_1(\omega r_2(n-j\kappa)/c)}{J_0(\omega r_2(n-j\kappa)/c)} \right] = y + zv + zwu \Im(\gamma - \gamma_0),$$

..... (29 a)

since u is a pure imaginary and v, ω, y, z are all real.

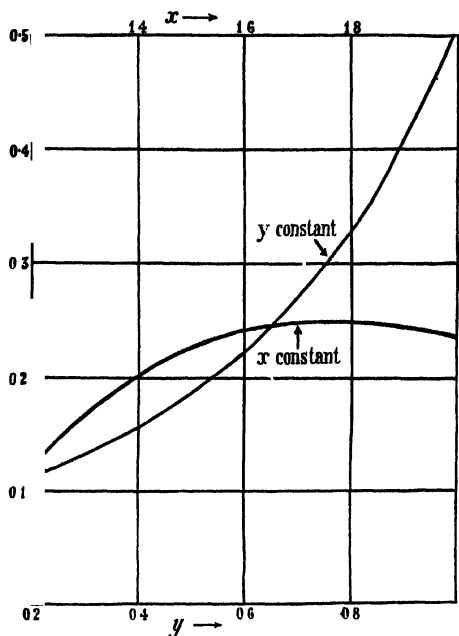


Fig. 4. Values of $\Im \left[\frac{1}{(x-jy)} \cdot \frac{J_1(x-jy)}{J_0(x-jy)} \right]$ for $x=1.6$ and y varying and for $y=0.5$ and x varying.

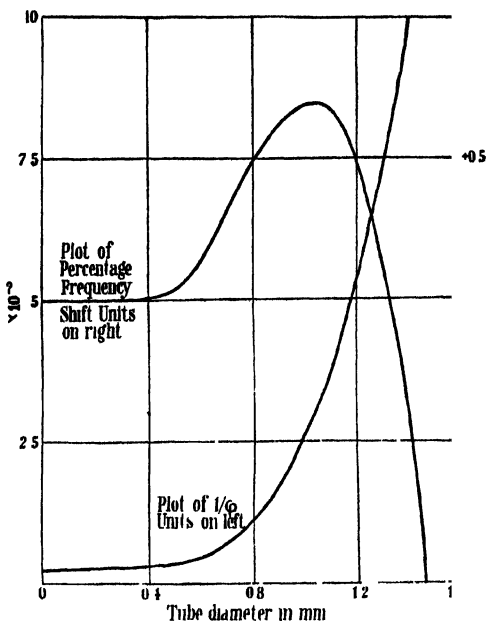


Figure 5 Approximate variation of $1/Q$ and percentage frequency shift with internal diameter of capillary, calculated and roughly checked experimentally for a typical polar liquid (water at 0°C , $n = 1.33$, $k = 2.8$).

Since the value of κ is known from an independent experiment, equation (29) can be solved numerically for n using tables of

$$\Im \frac{1}{z} \frac{J_1(z)}{J_0(z)} = \Im w(Q)$$

where $z = x - jy$.

If n were known independently, this could be used as a measurement of κ . The above function has been evaluated and is given in table 1, together with the real part of the function in table 2. The real part is included for information only, since it is not normally used, as the imaginary part of $Y - Y_0$, i.e. the frequency change of the resonator, cannot be determined experimentally with sufficient accuracy. This will be apparent from figure 5, where frequency shift and $1/Q$ are plotted against capillary tube radius, for a typical polar liquid. The error in calculating n from Y assuming that there is no frequency shift is in general

only a fraction of 1%, but originally a correction, computed from table 2, was included in the calculations. The behaviour of the frequency shift, which is in the normal direction for small tubes, and negative for large tubes, is interesting, and has been verified experimentally; it is probably due to the fact that the field

Table 1. Values of $\mu_m(Q)$

$y \rightarrow$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.5	0	0.006753	0.01347	0.01970	0.02587	0.03112					
0.6	0	0.008421	0.01685	0.02464	0.03198	0.03847	0.03847				
0.7	0	0.01038	0.02052	0.03041	0.03895	0.04665	0.05381	0.05974			
0.8	0	0.01251	0.02466	0.03615	0.04684	0.05615	0.06433	0.07100	0.07629		
0.9	0	0.01503	0.02959	0.04317	0.05579	0.06651	0.07612	0.08364	0.08983	0.0942	
1.0	0	0.01803	0.03526	0.05163	0.06627	0.07875	0.08958	0.09799	0.1046	0.1091	0.1122
1.1	0	0.02153	0.04237	0.06151	0.07846	0.09304	0.1051	0.1144	0.1210	0.1257	0.1284
1.2	0	0.02596	0.05085	0.07364	0.09353	0.1100	0.1232	0.1331	0.1402	0.1443	0.1494
1.3	0	0.03135	0.06154	0.08874	0.1116	0.1305	0.1447	0.1550	0.1615	0.1650	0.1661
1.4	0	0.03881	0.07530	0.0176	0.1343	0.1550	0.1703	0.1805	0.1861	0.1885	0.1878
1.5	0	0.04838	0.09326	0.1321	0.1631	0.1840	0.2012	0.2102	0.2141	0.2142	0.2117
1.6	0	0.06159	0.1178	0.1646	0.1998	0.2236	0.2382	0.2450	0.2462	0.2431	0.2374
1.7	0	0.08068	0.1524	0.2090	0.2480	0.2716	0.2831	0.2855	0.2821	0.2745	0.2646
1.8	0	0.1095	0.2027	0.2711	0.3118	0.3319	0.3368	0.3318	0.3216	0.3079	0.2932
1.9		0.1558	0.2801	0.3592	0.3969	0.4069	0.3998	0.3841	0.3642	0.3435	0.3220
2.0	0	0.2385	0.4063	0.4896	0.5107	0.4988	0.4715	0.4399	0.4074	0.3777	0.3500

never penetrates the polar liquid to a great depth, so that instead of the normal concentration of field in the dielectric, making the resonator effectively larger, the liquid actually contains less field than would be in its place in an empty resonator, so that the effective size is decreased.

Table 2. Values of $\Re(Q)$

γ	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
0.5	0.5163	0.5156	0.5131	0.5093	0.5040	0.4973					
0.6	0.5240	0.5233	0.5208	0.5160	0.5106	0.5028	0.4946				
0.7	0.5334	0.5322	0.5294	0.5245	0.5180	0.5098	0.5006	0.4902			
0.8	0.5448	0.5435	0.5402	0.5345	0.5274	0.5181	0.5074	0.4959	0.4838		
0.9	0.5585	0.5570	0.5530	0.5467	0.5380	0.5274	0.5154	0.5026	0.4887	0.4745	
1.0	0.5752	0.5735	0.5687	0.5611	0.5508	0.5384	0.5247	0.5099	0.4945	0.4788	0.4656
1.1	0.5950	0.5929	0.5870	0.5778	0.5656	0.5510	0.5349	0.5173	0.5000	0.4823	0.4660
1.2	0.6188	0.6163	0.6091	0.5975	0.5824	0.5650	0.5459	0.5260	0.5058	0.4855	0.4665
1.3	0.6476	0.6436	0.6349	0.6206	0.6020	0.5806	0.5577	0.5342	0.5111	0.4885	0.4669
1.4	0.6827	0.6785	0.6664	0.6479	0.6243	0.5980	0.5699	0.5424	0.5151	0.4899	0.4661
1.5	0.7267	0.7212	0.7050	0.6802	0.6497	0.6167	0.5824	0.5497	0.5183	0.4897	0.4640
1.6	0.7821	0.7742	0.7515	0.7178	0.6779	0.6357	0.5939	0.5546	0.5187	0.4871	0.4587
1.7	0.8539	0.8422	0.8097	0.7621	0.7086	0.6538	0.6024	0.5560	0.5155	0.4806	0.4500
1.8	0.9503	0.9318	0.8819	0.8149	0.7396	0.6686	0.6055	0.5518	0.5064	0.4687	0.4371
1.9	1.085	1.054	0.9731	0.8695	0.7656	0.6734	0.5990	0.5386	0.4900	0.4510	0.4194
2.0	1.288	1.229	1.0862	0.9231	0.7788	0.6652	0.5781	0.5136	0.4643	0.4260	0.3960

Normally, tube diameters > 0.6 mm. are not used, since the resonance curve is too flat to be measured without errors due to its asymmetry.

Tables 1 and 2 are evaluated from the identity (Watson)

$$J_n(x+y) = \sum_{m=0}^{\infty} J_m(x) J_{n-m}(y).$$

This leads to the following expansions:

$$J_0(x+jy) = J_0(x)I_0(y) - 2J_2(x)I_2(y) + 2J_4(x)I_4(y) \dots \\ + j[-2J_1(x)I_1(y) + 2J_3(x)I_3(y) \dots] \quad \dots \quad (30)$$

and

$$J_1(x+jy) = J_1(x)I_0(y) + J_1(x)I_2(y) - J_3(x)I_2(y) - J_3(x)I_4(y) \dots \\ + j[J_0(x)I_1(y) - J_2(x)I_1(y) - J_2(x)I_3(y) + J_4(x)I_3(y) \dots] \quad (31)$$

in which further terms can be neglected.

In figure 4 is shown the variation of $n(Q)$ with x and y , from which it will be seen that with x constant there is a portion of the curve over which the function hardly varies as y increases. If the capillary tube is selected with such a radius that the function falls on this part of the curve, a large inaccuracy (5%) in the independent value of $\kappa(\alpha y)$ will only result in a small inaccuracy ($\frac{1}{2}\%$) in $n(\alpha x)$.

§ 4. EXPERIMENTAL

The resonator, of internal diameter 1.912 cm., is drilled out of a solid brass cylinder, reamed and lapped to an accuracy of ± 0.02 mm. It is shown in figure 6. Since a very high Q is unnecessary, the surface need not be silvered; a Q of about 3000 was used throughout. The resonator is fed by two holes half a wavelength apart, 2.7 mm. in diameter, in the short side of a rectangular H_1 waveguide, which forms the centre part of a brass plate terminating the resonator at one end. Power is led from the resonator to the detector through another H_1 waveguide, which is fed from a small hole in the side of the resonator. The axis of this coupling hole makes an angle of 45° with the line of the two feeder holes so as to reduce the effect of unwanted modes. The resonator is tuned by a brass plunger fitting the barrel to 0.5 mm. clearance, and moved by a Moore and Wright micrometer head reading to 0.005 mm. with 2.5 cm. movement. The plunger is backed at its rim by attenuating plastic, which suppresses unwanted modes whose fields on the surface of the resonator are not zero. The glass capillary holding the liquid is a push fit in a hole drilled centrally in the plunger, and slides through two holes drilled centrally in the two short sides of the input waveguide, as the plunger is moved through the resonator. To eliminate dielectric pick-up from the waveguide the glass is

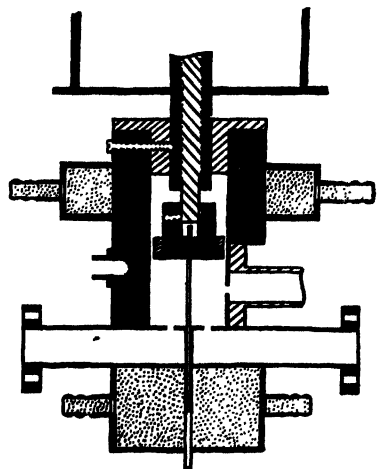


Figure 6. Resonator.

Note. For convenience the flat for the outlet hole shown normal to the waveguide and not at 45° .

screened by a fine metal tube where it crosses the guide. It is possible for the capillary tube to be central to 0.01 mm. at one end and 0.1 mm. at the other. The resonator is water-jacketed for temperature control both at the waveguide end, and at another point on its barrel. The temperature is read by a thermometer resting in, and making liquid contact with, a hole in the barrel.

The capillaries are drawn from Corning 707 low-loss glass whose dielectric constant is known at the frequency used. They are selected with an internal cross section uniform to one part in two hundred along the portion inside the resonator; this selection is done in the usual way by viewing a length of mercury at various points along the tube under a travelling microscope; the internal diameter is then measured to one part in five hundred by weighing a known length of mercury in the tube, at a known temperature. The walls of the tube should be about 0.05 mm. thick. The outside diameter should be uniform to one part in two hundred along its length, and the tube circular to one part in a hundred. It will be seen from the diagram of the resonator that there is sufficient room to allow an air space and seal at each end of the tube, but usually one end is not completely sealed, so that the water is always at atmospheric pressure.

The apparatus used for transmission and power measurement was similar to that described and illustrated in an earlier paper (Collie, 1946), with the addition of:

1. A piston attenuator which could be inserted into the waveguide instead of the resonator to calibrate the receiver.
2. A second H_{01} resonator connected as a monitor by a one-way feeder for watching frequency stability during a reading. Readings were rejected unless the frequency drift during a set of measurements was less than 50 kcs.
3. Further stabilization of the klystron power packs.
4. Screening round the transmitting tube of the type used in microwave signal generators.

Because of the design of resonator and the tight fitting of the capillary tubes in their plungers, the tubes were not removed for measurement of the Q of the empty resonator; from time to time the entire plunger was removed and replaced by a plain plunger for Q measurement; the change in frequency was measured by comparison with the monitoring resonator which had identical dimensions.

A typical Q curve and the linear plot obtained from it are shown in figure 2. The readings of receiver diode current against resonator tuning are taken for travel in both directions, and are converted to readings of power by use of the calibration curve of the receiver. The plot of inverse of power against square of breadth of Q curve should give a straight line from which the value of α is calculated as previously described. If a straight line is not obtained, it indicates that the resonance is unsatisfactory, or that the calibration of the receiver has changed, and this provides a useful check of each reading.

§ 5. RESULTS

The method described has been tried out on water at 21 °C. for which a value of $\kappa = 2.83$ is available (unpublished work of the authors). Three series of measurements using three different capillaries were made.

For each measurement, equation (29) has to be solved using the measured value of α and calculating the value of the constant $c = \pi w u$ from the measured values of r_0 , r_1 and r_2 and the known (Penrose, 1946) value of the dielectric constants of the glass.

The following results were obtained:

Table 3. Water at 21° C.

r_0 (mm.)	r_1 (mm.)	$\alpha 10^3$ (cm ⁻¹)	C	n
0.406	0.495	12.19	10.71	6.28
0.525	0.590	39.90	6.51	6.30
0.500	0.610	33.40	7.00	6.35

$$r_2 = 0.9560 \text{ cm.},$$

$$\epsilon' = 4.00,$$

$$\epsilon'' = 0.01. \quad \text{For Corning 707 low-loss glass.}$$

§ 6. DISCUSSION

Over the course of two years we have tried many methods of measuring the dielectric constant of water. The method described is the only one we have found for n at 1.25 cm. which is capable of $\pm 1\%$ accuracy, and a series of systematic measurements of the variation of n with temperature is being planned.

There is no doubt that if this work were being done again we should adopt the M.I.T. (Pound, 1946) system of frequency stabilization, which locks the oscillator on to a resonant cavity.

§ 7. ACKNOWLEDGMENTS

The work described in this paper has been carried out on behalf of the Director of Physical Research, Admiralty, and the authors wish to record their thanks for permission to publish. They thank Lord Cherwell for his continued interest and encouragement.

The suggestion of using an H_{01} resonator arose in a talk with Dr. L. A. P. Speirs, with whom we had many helpful discussions.

We are grateful to Mr. H. Ashcroft for the method of evaluating equations (30) and (31).

REFERENCES

- BETHE, 1942, "Perturbation theory of cavity resonators", *M.I.T. Report*.
 COLLIE, 1944, *Proc. Phys. Soc.*, **56**, 255.
 COLLIF, HASTED and RITSON, 1946, *Trans. Faraday Soc.* (Discussion on Dielectrics).
 PENROSE, 1946, *Trans. Faraday Soc.* (Discussion on Dielectrics).
 PINCHERLE, 1944, *Phys. Rev.*, **66**, 118.
 POUND, R. V., 1946, *Rev. Sci. Instrum.*, **17**, 490.
 WATSON, *Theory of Bessel functions* (Cambridge University Press), p.30.
 WILLIS JACKSON, 1941, *Ministry of Supply*, R. 287/Gen/35.

Pressure Broadening of the Inversion Spectrum of Ammonia: Part II—Disturbance of Thermal Equilibrium at Low Pressures

By B. BLEANEY AND R. P. PENROSE

Clarendon Laboratory, Oxford

MS. received 14 March 1947

ABSTRACT. The absorption at the centre of the line (3, 3) of the centimetre wavelength inversion spectrum of ammonia gas has been measured at pressures between 1.5 mm. and 0.01 mm. Hg. The absorption coefficient is constant at the higher pressures, as would be expected for a single line whose width is determined solely by pressure broadening. At the lower pressures the absorption coefficient falls by an amount dependent on the energy density in the resonator; this is due to the disturbance of thermal equilibrium in the gas through the absorption of energy from the radio-frequency field. A theory of the effect is developed from which the thermal relaxation time can be calculated by comparison with the experimental values of the absorption coefficient. It is found that this time is approximately 1.7_8 times greater than the mean time between collisions which interrupt the absorption (determined from the line breadth constant). This indicates that not every such collision is effective in restoring thermal equilibrium.

§1. INTRODUCTION

THE measurement of the absorption in ammonia gas at pressures of 1 cm. Hg to 60 cm. Hg has been described in a previous paper (Bleaney and Penrose, 1947 a). The observed values at pressures of 60 cm. and 10 cm. Hg were compared theoretically on the basis of the earlier analysis of the spectrum at pressures of about 1 mm. Hg (Bleaney and Penrose 1946 a, 1947 b). The observed and calculated values of the absorption at 10 cm. pressure were in close agreement over the whole range of the measurements ($0.63\text{--}0.92\text{ cm}^{-1}$), showing that the widths of the lines varied accurately at the pressure between 0.5 mm. Hg and 10 cm. Hg.

At pressures below 1 mm. Hg, the widths of most of the lines in this spectrum are small compared with the separation of the lines, and it is therefore possible to examine the lines individually. The width of the lines is still due almost entirely to collision broadening, since at a pressure of 0.5 mm. Hg the line breadth constants (equal to half the width of the line at half intensity) are found to lie between $2 \cdot 10^{-4}$ and $5 \cdot 10^{-4}\text{ cm}^{-1}$. Of other sources of broadening, that due to the Doppler effect was expected to be the most important; the line breadth constant associated with it is approximately 10^{-6} cm^{-1} , and its effect should not therefore become appreciable until pressures of about 10^{-3} mm. Hg are reached. Thus it was expected that a range of pressures varying by a factor of nearly 1000 should be available for the study of the effects of collision broadening on single lines.

In a region where the absorption is all due to a single spectral line whose width is determined entirely by collision broadening, the absorption at the centre of the line should be independent of the pressure. This is evident from the formula

$$\frac{\alpha}{\nu^2} = \frac{4\pi^2 |\mu_{mn}|^2 N_{mn}}{3kT \Delta\nu} \quad \dots\dots(1)$$

where α = absorption coefficient per cm. of path,

N_{min} = number of molecules per cc. in the lower and upper energy levels of the transition,

$\Delta\bar{\nu}$ = line breadth constant = half the width of the line at half intensity; since both N_{min} and $\Delta\bar{\nu}$ vary linearly with pressure, and the remaining constants are independent of the pressure. Since measurement of the absorption coefficient at the centre of the line is simpler and more accurate than a direct measurement of the width, it was decided to use this measurement as a test of the validity of equation (1). The line (3, 3) was chosen as the strongest and most convenient line for this purpose. Contrary to expectations, it was found that the absorption coefficient began to decrease when the pressure was about a tenth of a millimetre, and fell rapidly as the pressure was reduced still further. Investigation of the other lines of the spectrum showed that a similar effect occurred with all of them. The absorption at their centres began to fall at pressures of about 0.1 mm. and continued to decline as the pressure was reduced.

This fall in the absorption coefficient might be due to some other form of broadening of the line, which is independent of the pressure; the line breadth constant associated with this broadening must be of the order of 10^{-4} cm.⁻¹ to explain the observations. The following possibilities were considered:

(a) The Doppler width

Owing to the velocities of the molecules of a gas, a spectral line has a Doppler width given by the expression

$$\Delta\bar{\nu} = \left(\frac{\bar{v}}{c}\right) \sqrt{\frac{2RT}{M} \log_e 2}. \quad (\text{cm.}^{-1})$$

For ammonia, the value of $\Delta\bar{\nu}$ given by this is $1.2 \cdot 10^{-6}$ cm.⁻¹

(b) Interruption of the absorption owing to collisions with the walls.

In a container of volume V and wall area A , a molecule of mean velocity \bar{v} will make a collision with the wall $\frac{1}{4}A\bar{v}/V$ times a second. For the resonator used in this experiment, this will cause a broadening of the line of approximately

$$\Delta\bar{\nu} = \frac{1}{4}A\bar{v}/2\pi cV, \\ \therefore 3 \cdot 10^{-7} \text{ cm.}^{-1}$$

(c) The natural line breadth, due to spontaneous emission.

The line breadth constant associated with spontaneous emission is

$$\Delta\bar{\nu} = \frac{32\pi^3 \bar{\nu}^3 |\mu_{\text{min}}|^2}{3hc}. \quad (\text{cm.}^{-1}) \quad \dots\dots(2)$$

For wavelengths of about 1 cm. this gives a value of $\Delta\bar{\nu}$ of about 10^{-18} cm.⁻¹ In Einstein's theory of spontaneous emission, the density of radiant energy is assumed to be given by Planck's law; Purcell (1946) has, however, pointed out that with a resonator is associated one oscillator in a frequency range ν/Q , rather than $8\pi\nu^2/c^3$ per unit volume per unit band width, as with Planck radiation theory. Thus with a resonator Q of 10 000 the line breadth constant will be about 10^{-15} cm.⁻¹

(d) The natural line breadth due to stimulated emission and absorption.

A molecule in thermal equilibrium with its surroundings will undergo transitions owing to interaction with the radiation field. In the optical region the number of such transitions is small compared with the number of transitions due

to spontaneous emission, but the reverse is true at radio frequencies. The number n of transitions per second that a molecule undergoes in equilibrium in thermal radiation of energy density ρ is

$$n = \frac{8\pi^3 |\mu_{mn}|^2}{3h^2} \rho. \quad (\text{sec}^{-1}) \quad \dots\dots(3)$$

At these frequencies where $ch\bar{\nu} \ll kT$, this becomes

$$n = \frac{64\pi^4 |\mu_{mn}|^2 \bar{\nu}^2 kT}{3h^2 c} \quad (\text{sec}^{-1}) \quad \dots\dots(3a)$$

whence

$$\Delta\bar{\nu} = n/2\pi c = \frac{32\pi^3}{3h^2 c^2} |\mu_{mn}|^2 \bar{\nu}^2 kT. \quad (\text{cm.}^{-1}) \quad \dots\dots(3b)$$

This formula gives the width of a radio-frequency spectral line as observed in emission; for the inversion lines of ammonia this $\Delta\bar{\nu}$ is about $10^{-16} \text{ cm.}^{-1}$, or, if the emission excites a resonator tuned to the frequency of the line, $\Delta\bar{\nu}$ will again be about 1000 times greater. In absorption measurements the resonator is excited by radiation of an intensity immensely greater than that corresponding to room temperature, and the width of the line in absorption will be correspondingly greater. This point will be discussed later.

(e) Interruption of the inversion by a transition to another rotational level

Since the inversion frequency of the molecule depends strongly on the rotational state of the molecule, the absorption of radiation corresponding to a transition between the two levels of the inversion doublet will be interrupted if the molecule jumps to another rotational level. Transitions between the rotational levels will be caused by spontaneous emission, and by absorption and stimulated emission due to the thermal radiation. From equations (2) and (3) it can be seen that transitions between the rotational levels are considerably more frequent than similar transitions between the inversion sub-levels, since the energy density of black-body radiation is much greater at the higher frequencies of the rotational spectrum of ammonia. Calculation shows that the number of quanta absorbed per second by a molecule in the $J=3$ level causing a transition to the $J=4$ level is about 10 per second; the number of spontaneous and stimulated emissions causing transitions to the $J=2$ level are each of the same order. Hence the associated line breadth constant is about $10^{-10} \text{ cm.}^{-1}$

It is apparent that none of these phenomena is sufficient to cause a broadening comparable even with that due to the Doppler effect, except that due to the stimulated transitions considered under (d) above. This phenomenon cannot be considered as a simple broadening, but as a disturbance of thermal equilibrium in the gas. The absorption of energy from the radio-frequency field tends to equalize the populations of the two energy levels of the inversion doublet; this process is counteracted by the action of the collisions between the molecules, which tends to restore thermal equilibrium. As the pressure is reduced, the time between collisions gets longer, and the "relaxation time" for the return to thermal equilibrium increases. When the collision frequency becomes so low that the probability of making a transition in the interval between collisions approaches unity, the

average populations of the energy levels will no longer correspond to that appropriate to the "temperature" of the gas, but to some higher temperature. The intensity of the absorption will then decrease.

That this effect is of the correct order of magnitude to explain the observed diminution in the absorption coefficient is shown by the following calculation. If N_{JK} is the number of ammonia molecules per cc. in the rotation level characterized by the quantum numbers J, K , then the excess number in the lower level of the inversion doublet is $\frac{1}{2}N_{JK}(\hbar\nu_0/kT)$, where $\hbar\nu_0$ is the energy separation of the two levels of the doublet. To equalize the populations of the two levels, half of these must be transferred to the upper level, which requires an energy of $\hbar\nu_0 \cdot \frac{1}{4}N_{JK} \cdot (\hbar\nu_0/kT)$. This is the maximum energy that the gas can absorb from the radiation in a time (τ) equal to the mean interval between collisions.* Hence the power dissipated per cc. is

$$W = \frac{1}{4}N_{JK} \cdot (\hbar\nu_0)^2/kT. \quad \dots\dots(4)$$

For the line $(J, K) = (3, 3)$ at a pressure of 0.1 mm. Hg $(1/\tau) = 1.8 \cdot 10^7 \text{ sec.}^{-1}$, while $N_{JK} = 2.3 \cdot 10^{13}$; hence $W = 200 \text{ ergs/cc./sec.}$ Now the power dissipated in the resonator in these experiments is of the order of 1/10 milliwatt, or 1000 ergs/sec. At the centre of a strong line such as (3, 3) some 2/3 of this power would be absorbed by the ammonia, in a volume (allowing for the distribution of electric field in the resonator) of less than 2 cc., if the absorption has its normal intensity. Since this would require an absorption of energy by the ammonia greater than that given by equation (4), it is obvious that the strength of the absorption must be less than that corresponding to the maintenance of thermal equilibrium.

The treatment of the last paragraph leads to a simple expression of the condition that must be fulfilled if the true value of the absorption in a gas is to be measured. This condition is that the power absorbed by the gas must be small compared with the power given by equation (4). This equation contains two points of interest:

(a) W depends on the square of the resonant frequency of spectral line.

This arises because the energy absorbed in a transition varies with ν and so does the difference of population of the upper and lower levels of the transition.

(b) W varies with the square of the pressure in the gas.

This follows from the fact that both N_{JK} and $(1/\tau)$ vary linearly with the pressure.

The correctness of this explanation of the fall in the absorption coefficient at pressures of about 1/10 mm. Hg could quickly be verified experimentally by measurement with different power levels in the resonator. From (b) of the last paragraph it would be expected that the drop in the absorption coefficient would be the same under conditions such that the product (power dissipated in the resonator times square of the pressure of ammonia) is constant. This was found to be the case.

The phenomenon of the disturbance of thermal equilibrium in radio-frequency spectroscopy owing to the absorption of energy from the radiation was suggested and observed in experiments on the nuclear magnetic resonance by Purcell,

* For the purposes of this calculation it is assumed that the thermal relaxation time is identical with the mean time between collisions which interrupt the absorption of radiation.

Torrey and Pound (1946), and Bloch, Hansen and Packard (1946). In these experiments with liquid or solid samples, the thermal relaxation time is unknown, even as to order of magnitude, and a main interest lies in its determination by means of the decrease in the apparent absorption coefficient as the energy density of the radiation is increased. In the case of ammonia, the thermal relaxation time should be of the same order as the mean time between collisions* which is accurately known from the width of the spectral line, measured at low energy density. The opportunity arises, however, of determining whether every collision which is effective in interrupting the absorption of radiation is also effective in tending to restore thermal equilibrium. This may be accomplished by measurement of the thermal relaxation time by observation of the variation of the absorption coefficient with pressure at various power levels, which is more convenient experimentally than the complementary method of observing the decrease in the apparent absorption at a given pressure as the power level is increased. In the following sections the theory of the method is developed, and the predicted results are compared with those obtained experimentally by measurements at the centre of the line (3, 3). From this comparison the ratio of the thermal relaxation time to the mean time between collisions is obtained.

§ 2. THEORY

Since the strength of the absorption depends only on the *difference* between the number of molecules occupying the lower and the number occupying the higher of the two energy levels between which transitions are taking place, it is sufficient to work in terms of this difference, rather than the actual numbers themselves. Let n_0 be the excess number of molecules in the lower state corresponding to the actual temperature of the gas, and let n be the actual excess when radio-frequency power is applied. Then in the equilibrium condition the rate at which the r.f. power is causing n to decrease must be balanced by the action of collisions which tend to restore n to the value n_0 corresponding to thermal equilibrium. If β is the number of transitions per second per molecule caused by the radiation, then the rate of change of n is $(-2\beta n)$, since the transfer of one molecule from the lower to the upper state reduces the excess by two. The rate of change of n due to collisions may be written (cf. Fröhlich, 1946) as $(n_0 - n)/\tau_c$, where τ_c is the mean time between collisions that are effective in tending to restore thermal equilibrium. Hence we have

$$\frac{dn}{dt} = -2\beta n + (n_0 - n)/\tau_c. \quad \dots\dots(5)$$

In the steady state $dn/dt = 0$; hence

$$n/n_0 = 1/(1 + 2\beta\tau_c) \quad \dots\dots(5a)$$

and this equals the ratio of the observed absorption coefficient (α) to the true absorption coefficient (α_0) which would be observed if the energy density of the radiation were infinitely small.

Since the ratio n/n_0 is only appreciably different from unity for very low pressures, where the spectral lines are narrow, it is sufficiently accurate in the

* In this paper, where the word "collisions" occurs unqualified, it is to be understood as referring to collisions which produce interruption of the absorption of radiation. This restriction is introduced to avoid constant repetition of this phrase.

evaluation of β to neglect the second-order term introduced by Van Vleck and Weisskopf (1945) and Fröhlich (1946) into the collision broadening theory of Lorentz. Then at a place where the electric field of the radiation is $E \cos 2\pi\nu t$ the chance of a molecule making a transition between two energy levels W_m and W_n in a time t is

$$(a_m^* a_m) = |\mu_{mn}|^2 E^2 \frac{\sin^2 \left\{ \frac{(W_m - W_n - \hbar\nu)t}{\hbar} \right\}}{(W_m - W_n - \hbar\nu)^2}$$

(cf. Pauling and Wilson, *Introduction to Quantum Mechanics*, Chap. XI, equation (40-12)). Here

$$\mu_{mn} = \int \psi_m^* \mu \psi_n d\sigma$$

where μ , is the component of the electric dipole moment in the direction of the electric field of the radiation. Since the dipoles are randomly oriented,

$$(\mu_{mn})^2 = (\mu_{ymn})^2 = (\mu_{zmn})^2 = \frac{1}{3}(\mu_{mn})^2.$$

In the measurements to be described the frequency of the applied radiation corresponded to the centre of the spectral line; i. e. $\hbar\nu = W_m - W_n$ and (6) becomes

$$(a_m^* a_m) = \frac{|\mu_{mn}|^2}{\hbar^2} \frac{t^2}{3} E^2. \quad \dots\dots (6a)$$

The value of t in this equation is that corresponding to the actual time between successive collisions; since this time is random, an average must be taken over all possible values. For present purposes it is sufficient to assume an exponential distribution $(1/\tau) e^{-t/\tau}$ of the values of t . Then, since $\overline{t^2} = 2\tau^2$, (6a) becomes

$$(\overline{a_m^* a_m}) = \frac{2\pi^2 |\mu_{mn}|^2}{\hbar^2} \frac{E^2 \tau^2}{3} \quad \dots\dots (6b)$$

and the β of equation (5) equals $(\overline{a_m^* a_m})/\tau$.

The appropriate value of τ in equation (6b) is the mean time between collisions which are effective in interrupting the absorption of radiation, and τ is therefore related to the line breadth constant for collision broadening. The value of τ may therefore be found by observation of the width of the spectral line under conditions where broadening due to causes other than collisions between the molecules is negligible.

In applying equation (6b) to a case where the absorption of energy from the radiation is so great that thermal equilibrium is disturbed, it is necessary to examine its validity under these conditions. The expression for the transition probability (equation (6)) is derived by means of first-order perturbation theory, where the assumption is made that the chance of a transition in the time t is very small. It is obvious from equation (5a), however, that n/n_0 will only be materially smaller than unity when the term $2\beta\tau_e$ is no longer small. If τ_e is approximately equal to τ , the value of $(\overline{a_m^* a_m})$ must be of the same order as $\beta\tau_e$; that is, unity or more. Thus equation (6b) is here being applied under conditions where its method of derivation is no longer justified, if n is materially less than n_0 . However, in the absence of an exact theory* it is interesting to examine whether the variation of the absorption

* Professor M. H. L. Pryce informs us that preliminary work on a theory which avoids these limitations indicates that the formula (equation (7)) for (n/n_0) is valid even when the apparent absorption coefficient is materially smaller than the true value.

coefficient observed experimentally can be adequately represented by formulae deduced in this way.

If the absorption is observed using a plane wave traversing the unbounded medium, equation (5) becomes simply

$$\alpha/\alpha_0 = n/n_0 = 1 / \left(1 + \frac{4\pi^2}{3\hbar^2} |\mu_{mn}|^2 E^2 \tau \tau_c \right) \quad \dots\dots (7)$$

$$= 1 + \frac{32\pi^3}{3\hbar^2 c} |\mu_{mn}|^2 W \tau \tau_c \quad \dots\dots (7a)$$

where $W = cE^2/8\pi =$ energy transmitted across unit area per second.

In a waveguide, or cavity resonator, the effect is more complicated owing to the inhomogeneous distribution of the electric field. Most of the absorption occurs at a place where the electric field is strongest, and at this point the value of n/n_0 will fall below unity by the greatest amount. At other places the disturbance of thermal equilibrium owing to the effect of the radiation will be less, and the values of n/n_0 will be closer to unity. Clearly the observed value of α/α_0 will be that corresponding to an average value of n/n_0 , obtained by evaluation of the integral

$$\alpha/\alpha_0 = \frac{\int \{E^2/(1 + 2\beta\tau_c)\} dV}{\int E^2 dV} \quad \dots\dots (8)$$

the integral being taken over the whole volume in which radiation is acting on the gas. An explicit expression for this integral cannot be given for the general case, but it may be shown that for a given shape of cavity and mode of excitation its value depends only on the non-dimensional parameter

$$a' = (4\pi^2/3\hbar^2) \cdot |\mu_{mn}|^2 \cdot \tau \tau_c \cdot (gW'V),$$

where $W' =$ total energy stored in cavity,

$V =$ total volume of cavity,

$g =$ dimensionless constant calculated from the configuration of the electric field.

In these experiments a cylindrical cavity in the H_{10} mode was used, and the relation between (α/α_0) and a' is shown graphically in figure 1. The evaluation of the integral in equation (8) for this case is given in Appendix A.

For the comparison of this theory with the experimental results the values of the variables in the parameter (a') are required. For the inversion lines of ammonia the matrix element for a rotational state denoted by the quantum numbers (J, K) is

$$|\mu_{JK}|^2 = \mu^2 \cdot K^2/(J^2 + J).$$

The remaining quantities can only be determined by experiment. The mean time

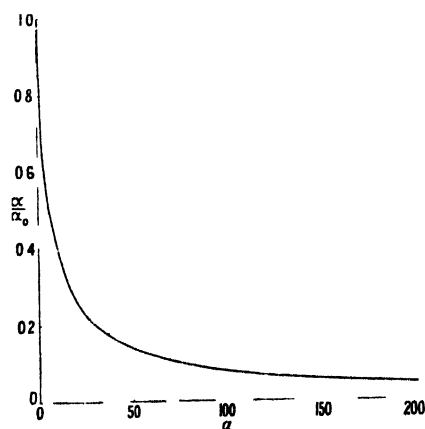


Figure 1 Fractional drop in absorption (α/α_0) as a function of power level in an H_{10} cylindrical resonator

between collisions which interrupt the absorption of radiation (τ) is found from the line breadth constant ($\Delta\bar{\nu}$) measured under conditions where only collisions are significant in broadening the line, using the relation

$$\tau = 1/(2\pi\Delta\bar{\nu}c).$$

The total energy W' stored in the resonator can be found by measuring the power dissipated in the resonator and the circuit magnification factor Q ; the volume V is known from the dimensions of the cavity.

There remains the thermal relaxation time τ_c which cannot be determined independently. Since, however, the thermal relaxation is due to collisions between the molecules, one would expect that τ_c should vary inversely with the pressure in the same way as the collision time τ . It is therefore convenient to work in terms of the ratio of these two times (τ_c/τ) as a parameter whose value should be independent of the pressure. The value of a' may then be written

$$a' = \frac{|\mu_{JK}|^2 (\tau_c/\tau)}{3(\hbar c)^2 \Delta\bar{\nu}^2} \left(\frac{gW'}{V} \right)$$

where only the ratio (τ_c/τ) is undetermined.

The development of the theory has thus been carried as far as possible, and its results can be subjected to experimental test. The purpose of such a test is two-fold:

- (1) To examine whether the formulae derived by this theory are applicable when thermal equilibrium is seriously disturbed, and the ratio of the apparent to the true absorption coefficient is very much less than unity.
- (2) To determine the ratio of the thermal relaxation time τ_c to the mean time between collisions τ .

If it is found possible to represent the whole range of the experimental results by curves calculated from the theory with a single value of the undetermined parameter (τ_c/τ), it seems justifiable to assume that the theory is satisfactory and that the results yield a value of the thermal relaxation time.

§ 3. EXPERIMENTAL APPARATUS

The apparatus used in these experiments was the same as that previously described (Bleaney and Penrose 1947 b) for measurements on the spectrum of ammonia gas at pressures of about 1 mm. Hg, and only a brief outline need be given here. The lay-out of the apparatus can be seen from the photograph (figure 2). Power from a reflexion klystron A is fed into a rectangular waveguide, its magnitude being adjusted by means of the variable attenuator C. Its wavelength can be determined by means of the sharply tuned wavemeter E with an accuracy of one or two parts in 10 000. The experimental chamber H is a cylindrical cavity resonant in the H_0 mode, which can be evacuated or filled with ammonia at any desired pressure. The cavity is tuned by a plunger driven by a micrometer head M, whose motion is magnified, when necessary, by the mirror, lamp and scale.

The cavity is excited by power fed from the waveguide through two small holes in the narrow side of the rectangular guide. The waveguide is terminated by an attenuator J, followed by a silicon-tungsten crystal rectifier (not visible) which acts as a monitor. To measure the power flowing along the guide, this can be replaced by a thermistor bolometer. Resonance in the cavity is detected by

another crystal rectifier K, preceded by an attenuator L, coupled to the cavity through a small hole in the side of H. The rectified current from this crystal is proportional to the energy density in the cavity, and the strength of the absorption of the ammonia is measured by means of the change in the energy density in the cavity on admitting the ammonia. If δ_0, δ_1 are the crystal current readings before and after admitting the ammonia, then the absorption coefficient per cm., α , may be calculated from the relation

$$\alpha = \frac{2\pi}{\lambda Q_0} \left\{ \sqrt{\frac{\delta_0}{\delta_1}} - 1 \right\}, \quad \dots\dots (12)$$

where λ is the wave-length of the radiation in free-space and Q_0 is the circuit magnification factor of the empty resonator. Q_0 is determined from the movement of the micrometer head required to detune the resonator to the half-power points, and can be measured with an accuracy of a few per cent.

Measurements

To test the theory of the disturbance of thermal equilibrium developed above, the procedure adopted was to measure the apparent absorption at the centre of the line (3, 3) at pressures from 1.5 mm. Hg down to a few hundredths of a mm., using two different power levels. This line is one of the strongest in the ammonia spectrum, and at these pressures the absorption due to the tails of neighbouring lines is very small. The centre of this line lies at 0.7964 cm^{-1} , and can be determined very accurately by locating the maximum of the absorption at a pressure of about 0.1 mm. The reading of the micrometer head attached to the resonant cavity corresponding to this maximum was observed in a preliminary experiment, and the subsequent measurements were all made with the frequency of the oscillator adjusted to be within 0.00002 cm^{-1} of the exact centre of the line.

The amount of r.f. power entering the cavity cannot be measured directly, but may be deduced from measurements of the power reaching a detector placed at the end of the input waveguide when the cavity is first detuned from, and then tuned to, resonance. A simple calculation based on the assumption that both the generator and the detector are matched to the waveguide shows that if the drop in the detector reading is a small fraction r of its initial reading, then the power entering the resonator is a fraction $r/(1+r)$ of that incident on the detector before tuning the cavity. The power flowing along the main waveguide was measured by a bolometer whose sensitive element was a thermistor. On adjusting the cavity to resonance the reading of the bolometer fell by 13% and the power entering the cavity is therefore a little under 12% of the incident power. The remaining 1% is reflected back to the generator owing to the slight mismatch caused by the impedance reflected into the guide when the cavity is tuned.

At the lowest power level the thermistor was not sufficiently sensitive to measure accurately the drop in power on tuning the cavity as the power incident on the thermistor could only be determined to within 10%. To verify that the fraction of this power entering the cavity was the same as at the higher level, the thermistor was removed, and the attenuator J followed by a crystal detector was replaced. It was found that the drop in the rectified current from the crystal on tuning the cavity was in good agreement with that obtained with the thermistor at the higher power level.

These methods were used to determine the power dissipated, and hence the energy density in the resonant cavity when it was evacuated. When ammonia is admitted the Q of the cavity drops owing to the absorption of energy by the gas, and less power is drawn by the cavity from the main waveguide. The energy stored in the cavity changes, therefore, with the square of Q . It is not necessary to determine the Q and the power entering the cavity at each pressure of ammonia, however, since the change in the energy density in the cavity is measured directly by the change in the power registered by the crystal detector K. From the latter the absorption coefficient of the ammonia is also calculated by equation (12).

The power output from the oscillator to the waveguide remained constant to within 2%, during the experiments. The procedure was therefore as follows: After the measurement of the power entering the evacuated resonator the rectified current of about 1 μ a. from the crystal K was observed by a sensitive galvanometer. Ammonia was then admitted to the cavity, its pressure being measured by the McLeod gauge shown in figure 2, and the new reading of the galvanometer observed. The cavity was evacuated and the procedure repeated with different pressures of ammonia until sufficient points were obtained to delineate a curve of absorption coefficient against pressure.

8.4 RESULTS

Measurements were made at two power levels differing by a factor of 70. At the higher level the power in the waveguide was 12.3 mw. at the coupling to the resonant cavity; this was the maximum that could be obtained using sufficient attenuation in the guide to prevent errors due to change of oscillator frequency on tuning the resonator. The power entering the resonant cavity when evacuated was 1.4 mw.

In the second series of experiments, the power was reduced to as low a value as possible, the limit being set by the sensitivity of the detector. The power level in the waveguide was 0.16 mw., and the power entering the resonator was approximately 20 μ w.

The resonator was used at its second resonance and was 1.82 cm. long; its radius was 1.05₁ cm. The circuit magnification factor Q_0 of the empty resonator was measured and found to be 8500.

The change of the absorption coefficient with pressure at the two power levels is shown in figure 3, where the absorption is expressed both per cm. of path and in decibels per kilometre. It will be seen that the drop in the apparent absorption as the pressure is reduced occurs at much greater pressures when the power level is higher. In fact the pressures in the two curves at which the absorption has fallen to a given value are very closely in the ratio of $\sqrt{70}$:1 i.e. the square root of the ratio of the two powers, as would be expected if the phenomenon is due to disturbance of thermal equilibrium. At pressures above 0.2 mm. Hg the absorption coefficient is very nearly constant when the lower power level is used, showing that contributions to the width of the line from causes other than collision broadening are negligible. The experimental values of the absorption coefficients show a slight increase at the higher pressures, due to the contributions from neighbouring lines. The additional absorption from this cause has been calculated, and subtracted from the experimental points before plotting them in figure 3.

To compare these results with the variation of absorption with pressure and power level predicted by the theory developed earlier in this paper, the width of the spectral line is required, measured at low energy density so that thermal equilibrium is not disturbed. As reported previously (Bleaney and Penrose 1947 b) the width is found from the shape of the line at a pressure of 0.5 mm. Hg, determined by measurement of the absorption coefficient at a sufficient number of frequencies to delineate the line. These measurements are shown in figure 4, where the absorption coefficient is plotted against the wavelength in the resonant cavity, small

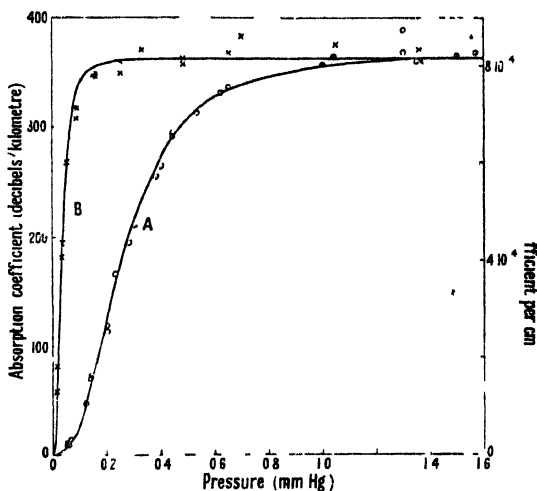


Figure 3. Variation of absorption with pressure at the centre of the line $(J, K) = (3, 3)$ for two different power levels.

- A. Power entering evacuated cavity - 12.3 milliwatts
 - Experimental points
 - Theoretical curve, assuming $(\tau_e/\tau) = 1.75$
- B. Power entering evacuated cavity - 20 microwatts
 - x Experimental points.
 - Theoretical curve, assuming $(\tau_e/\tau) = 1.75$.

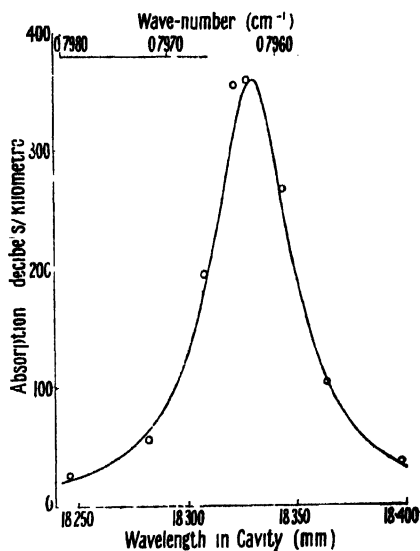


Figure 4. The line $(J, K) = (3, 3)$ at a pressure of 0.5 mm Hg.

changes in which are proportional to changes in the wave-number. The full line shows the theoretical absorption curve, calculated from the formula

$$\alpha_0 = \frac{4\pi^2 |\mu_{JK}|^2 \bar{\nu}^2 N_{JK}}{3kT} \left\{ \frac{\Delta \bar{\nu}}{\Delta \bar{\nu}^2 + (\bar{\nu}_0 - \bar{\nu})^2} \right\}$$

where

$$|\mu_{JK}|^2 = |\mu_{3,3}|^2 = \frac{3}{4} (1.44 \cdot 10^{-18})^2 \text{ e.s.u.}$$

The value $4.5 \cdot 10^{-4} \text{ cm}^{-1}$ for the line breadth constant $\Delta \bar{\nu}$ has been taken as giving the best fit with the experimental points. The agreement between the experimental and theoretical values is very close, and the error in the value of $\Delta \bar{\nu}$ cannot be more than 1 or 2%. Since $\Delta \bar{\nu}$ varies linearly with the pressure, one may write

$$\Delta \bar{\nu} = 9.0 \cdot 10^{-4} p_{\text{mm}} \quad (\text{cm}^{-1})$$

An expression relating the constant α' of equation (11) with the experimental data can now be found. The energy stored in the resonant cavity is $W' = QW/\omega$. Here the measured quantities are not W and Q but W_0 , the power dissipated in the

empty resonator, whose circuit magnification is Q_0 . When ammonia is admitted so that the rectified current from the crystal detector K falls from δ_0 for the evacuated resonator to δ_1 , the value of W' is

$$W' = (Q_0 W_0 / \omega)(\delta_1 / \delta_0).$$

From these equations, and data given earlier in this section, the value of a' is found to be

$$a' = (0.63 p_{\text{mm}}^2)(\delta_1 / \delta_0)$$

at the higher power level, and

$$a' = (0.0090 p_{\text{mm}}^2)(\delta_1 / \delta_0)$$

at the lower power level.

The values of (δ_1 / δ_0) corresponding to each point in figure 3 are known from the measurement of the absorption. As a first attempt, the theoretical drop in the absorption coefficient was calculated assuming that the thermal relaxation time τ_c was equal to the mean time between collisions. The curves of (α, p) constructed in this way were found to be similar to the experimental curves, but the decrease of the apparent absorption was rather less than that observed experimentally. For both power levels the discrepancy was such that it appeared likely that the introduction of a simple multiplying factor of rather less than 2 into the values of a' would shift the calculated curves for both power levels into good agreement with the experimental points. To find the best value of this multiplying factor, the values of the constant a' which would be required to give the observed drop in the absorption coefficient were calculated for each of the experimental points at the higher power level. It was found that the values of a' for the most reliable points were between 1.7 and 1.8 times greater than those calculated on the assumption $(\tau_c / \tau) = 1$. The theoretical curves for both power levels were therefore recalculated assuming $(\tau_c / \tau) = 1.7$, and are shown as full lines in figure 3. It will be seen that the experimental results are in good agreement with the theoretical predictions made on this basis.

§ 5. DISCUSSION

Although the absorption coefficient has fallen at the lower pressures to about one-twentieth of the value which would be observed if the energy density of the radiation were sufficiently small, the theory is still successful in representing the experimental results. It may therefore be inferred that the formula (7) for the apparent diminution in the absorption is valid even under these conditions where its method of derivation cannot be justified. The success of the theory makes it permissible to discuss the significance of the thermal relaxation time found by means of these experiments.

A few simple considerations indicate that the value of the ratio of the thermal relaxation time (τ_c) to the mean time (τ) between collisions which are effective in interrupting the absorption of radiation should lie between fairly narrow limits. Since it seems impossible to conceive of a collision mechanism which is effective in producing thermal relaxation without interrupting the absorption of radiation, it is reasonable to suppose that (τ_c / τ) cannot be less than unity. On the other hand a "kinetic theory collision", i. e. a collision in which there is an appreciable transfer of momentum between the molecules, must surely be effective in producing thermal relaxation. This leads to an upper limit for (τ_c / τ) . The value found for

(τ_e/τ) in these experiments by no means approaches this upper limit, however. Viscosity measurements show that an appreciable transfer of momentum between ammonia molecules only occurs when they approach to within a distance of about 3 Å., whereas the value of the line breadth constant for this absorption shows that the interaction of the molecule with the radiation field is interrupted by the approach of another molecule to a distance of about 14 Å.

Consideration of the experimental errors suggests that the value 1.7₅ of the ratio (τ_e/τ) assumed in fitting the theoretical to the experimental values is probably correct to within $\pm 10\%$. The distance between the molecules at which the interaction is sufficient for the purposes of thermal equilibrium is thus about $14/\sqrt{1.75} = 11$ Å., which is still considerably greater than the kinetic-theory diameter of the molecule. It must be remembered, however, that the absorption only effects the distribution of molecules amongst the two energy levels of the inversion doublet whose separation is very small (0.8 cm^{-1}). The restoration of the correct thermal distribution may therefore be effected by the interaction of two molecules at a distance much greater than that required for the transfer of momentum.

A lengthy discussion of the significance of the difference between the thermal relaxation time and the mean time between collisions which interrupt the absorption of radiation would seem to be premature. The theory of collision broadening of Van Vleck and Weisskopf is based on the concept of a classical harmonic oscillator whose transient amplitude and phase after a collision are distributed in accordance with a Boltzmann function. Thus every collision is effective in restoring thermal equilibrium. On the other hand, the concept of the disturbance of thermal equilibrium is based on the existence of discrete energy levels, and the theory is essentially quantum-mechanical, as is obvious from the presence of Planck's constant in the formulae. It is thus very difficult to discuss the phenomenon from the point of view of the theory of Van Vleck and Weisskopf, and no analogous theory has yet been developed on quantum-mechanical principles. It seems pertinent to enquire, however, what meaning can be attached to a "collision" on such a theory. The absorption of energy from the radiation tends to equalize the numbers of molecules in the upper and lower levels between which transitions are taking place, and thermal relaxation tends to restore the unequal distribution appropriate to the temperature of the gas. That is, thermal relaxation collisions cause a net transfer of molecules from the upper to the lower level. The fact that τ is less than τ_e shows, however, that a large proportion of the collisions which interrupt the absorption are not effective in causing such a transfer. The fact that they cause a broadening of the spectral line shows that they must effect the wave function of the molecule in some way, even though they leave the molecule in the same energy level. It may prove necessary, therefore, to consider the phase of the wave function (which is ordinarily neglected), and to attribute to a collision a random change of phase in the wave function similar to the change of phase of the harmonic oscillator in the classical theory. Full elucidation of this point must await the proper formulation of a wave-mechanical theory of collision broadening.

§ 6. CONCLUSION

Apart from its intrinsic interest in the study of collision broadening, the phenomenon described in this paper is important in connection with the limitation of the

resolution of close spectral lines which it causes. This limitation arises in a two-fold manner; the absorption is reduced, making the lines more difficult to observe; but it is reduced less in the wings of a line than in the centre, giving the line an artificial breadth. The results of this paper support the assumption made in the theory that the line breadth due to collisions varies directly as the pressure, at any rate down to 0.03 mm. Hg. There is no reason to believe that this law will not be obeyed down to pressures of about 0.001 mm. Hg where the Doppler breadth would be of the same order as the collision breadth. To avoid disturbance of thermal equilibrium at such a pressure, the power entering a resonator, such as used in our experiments, would have to be reduced to about 10^{-10} watt. The power reaching a loosely coupled crystal detector would be about 10^{-12} watt, which is well above noise level if a superheterodyne method of detection is used. It should thus be possible to observe lines whose widths are about 10^{-6} cm.¹

The phenomenon of disturbance of thermal equilibrium is not confined to the resonant-cavity method of measuring absorption, though it occurs at lower powers and higher pressures than in a waveguide transmission method owing to the higher energy density in the resonant cavity. The decrease in intensity of the ammonia lines observed by Good (1946) at low pressures in a waveguide has been attributed by the authors (Bleaney and Penrose 1946b) to disturbance of thermal equilibrium.

The derivation of the formula quoted there is given in Appendix B; the assumption was made, however, that the thermal relaxation time is equal to the mean time between collisions. The fact that it is found experimentally to be somewhat longer shows that the effect on the absorption will be rather more pronounced than calculated by the authors in that paper. The phenomenon of the disturbance of thermal equilibrium in waveguide measurements has been confirmed independently by Townes (1946), who shows that the effect is of the right order of magnitude if the thermal relaxation time is approximately equal to the collision time.

ACKNOWLEDGMENTS

The authors are indebted to Professor M. H. L. Pryce for considerable discussion and detailed criticism of the theory of this paper.

The work described in this paper was carried out partly on behalf of the Director of Scientific Research, Admiralty, and the authors wish to record their thanks for permission to publish. They also acknowledge gratefully a Senior Studentship from the Royal Commission for the Exhibition of 1851.

APPENDIX A

Evaluation of $\langle \alpha/\alpha_0 \rangle$ for a cylindrical H_0 resonator

In these experiments, a cylindrical cavity resonator was used excited in the H_0 mode, and the azimuthal component of the electric field at the point whose cylindrical coordinates are (r, ϕ, z) is

$$E = E_0 J_1(kr) \sin(2\pi z/\lambda_g), \quad \dots\dots (9)$$

where k is determined from the relation $(kb) = 3.832$, b being the radius of the cylinder, and λ_g is the wavelength in the cavity. The other components of the electric field are zero.

Insertion of (9) into (8) yields a double integral, to be evaluated over the coordinates r and z . The latter integration is simpler, being of the form

$$\overline{(n/n_0)} = A \int_0^\pi \{\sin^2 \theta / (1 + a \sin^2 \theta)\} d\theta \quad \dots\dots (10a)$$

where

$$a = \frac{4\pi^2}{3\hbar^2} |\mu_{mn}|^2 E_0^2 J_1^2(kr) \tau \tau_c.$$

This integral may be evaluated explicitly, giving

$$\overline{(n/n_0)} = \frac{2}{a} \left(1 - \frac{1}{\sqrt{1+a}} \right), \quad \dots\dots (10b)$$

where the constant A has been determined by letting a tend to zero.

The remainder of the double integral is of the form

$$\overline{(n'/n_0)} = A' \int_0^b \frac{2\pi r J_1^2(kr)}{J_1^2(kr)} \left(1 - \frac{1}{\sqrt{1+a'J_1^2(kr)}} \right) dr,$$

where

$$a' = \frac{4\pi^2}{3\hbar^2} |\mu_{mn}|^2 E_0^2 \tau \tau_c.$$

Here the $2\pi r$ arises from the element of volume, and the $J_1^2(kr)$ in the numerator from the variation of the electric field strength in the numerator of equation (8). The integral may be written

$$\overline{(n'/n_0)} = A'' \int_0^{0.832} x \left(\frac{1}{1 - \sqrt{1+a'J_1^2(x)}} \right) dx.$$

The constant A'' is determined by the relation (obtained by letting $a' \rightarrow 0$)

$$\begin{aligned} 1 &= A'' \int_0^{0.832} x \frac{a'}{2} J_1^2(x) dx \\ &= A'' a' 0.60. \end{aligned}$$

Hence

$$\frac{\alpha}{\alpha_0} = \overline{(n'/n_0)} = \frac{1.67}{a'} \int_0^{0.832} x \left(1 - \frac{1}{\sqrt{1+a'J_1^2(x)}} \right) dx. \quad \dots\dots (11)$$

There does not seem to be any method of evaluating this integral except by numerical computation for different values of the parameter a' . This has been carried out for values of a' up to 200, at which the apparent absorption coefficient is smaller than the true value by a factor of about 20. The results are shown in figure 1, where (α/α_0) is plotted as ordinate and a' as abscissa.

It is convenient to relate a' to the total energy W' stored in the cavity. For a cylindrical H_0 resonator, whose length is m half-wavelengths, we have

$$\begin{aligned} W' &= \frac{1}{8\pi} \int_0^{\frac{m\lambda_g}{2}} E_0^2 \sin^2 \frac{2\pi z}{\lambda_g} dz \int_0^b 2\pi r J_1^2(kr) dr \\ &= E_0^2 V \cdot 3.2 \cdot 10^{-3}, \end{aligned}$$

where $V = \pi b^2 \frac{m\lambda_g}{2}$ = volume of cavity.

Hence

$$a' = \frac{4\pi^2}{3\hbar^2} |\mu_{mn}|^2 \tau \tau_c \left(\frac{3.1 \cdot 10^3 W'}{V} \right).$$

It is obvious that similar computations may be carried out for any shape of resonator and mode of excitation, and that the results may be expressed in the form of a curve of (α/α_0) (similar to, though not identical with, that shown in figure 1) against a non-dimensional parameter a' which may be written

$$a' = \frac{4\pi^2}{3\hbar^2} |\mu_{mn}|^2 \tau \tau_c \left(\frac{gW'}{I'} \right),$$

where g is a dimensionless constant which can be calculated from the field configuration in the cavity.

APPENDIX B

Evaluation of (α/α_0) for rectangular H_{11} waveguide

In a rectangular waveguide excited in the H_{11} mode the electric field has only one component, parallel to the narrow side, which varies as

$$E = E_0 \sin \pi x/b$$

from side to side, where b is the wide dimension of the guide. This leads to an integral similar to (10a) above, with a solution

$$\frac{\alpha}{\alpha_0} = \overline{n/n_0} = \frac{2}{a} \left(1 - \frac{1}{\sqrt{1+a}} \right),$$

where

$$a = \frac{4\pi^2}{3\hbar^2} |\mu_{mn}|^2 \tau \tau_c E_0^2.$$

The power flowing down the guide is

$$W = \frac{\lambda}{\lambda_g} \frac{A c E_0^2}{2 \cdot 8\pi},$$

where A = cross-sectional area of guide,

λ = wavelength in free space,

λ_g = wavelength in guide.

Hence, substituting for τ in terms of $\Delta\bar{\nu}$, a becomes

$$a = \frac{16\pi}{3(\hbar c)^2} \frac{|\mu_{mn}|^2}{(\Delta\bar{\nu})^2} \left(\frac{W \lambda_g}{A c \lambda} \right) \left(\frac{\tau_c}{\tau} \right).$$

The equation for (α/α_0) can only be applied to a short length of guide, where W is substantially constant. If a long length of guide is used, in which there is considerable attenuation due to the absorption of radiation by the gas, a further averaging of (n/n_0) over the length of the guide must be carried out. No explicit expression can be given for this case, because the power will not vary exponentially along the guide when thermal equilibrium is disturbed. The attenuation will, in fact, be less near the input, where the energy density is greater, than near the output.

REFERENCES

- BLEANEY and PENROSE, 1946 a, *Nature, Lond.*, **157**, 339; 1946 b, *Phys. Rev.*, **70**, 775; 1947 a, *Proc. Phys. Soc.*, **59**, 418; 1947 b, *Proc. Roy. Soc., A*, **189**, 358.
 BLOCK, HANSEN and PACKARD, 1946, *Phys. Rev.*, **70**, 127.
 FRÖHLICH, 1946, *Nature, Lond.*, **157**, 478.
 GOOD, 1946, *Phys. Rev.*, **70**, 213.
 PURCELL, 1946, *Phys. Rev.*, **69**, 681.
 PURCELL, TORREY and POUND, 1946, *Phys. Rev.*, **69**, 37.
 TOWNES, 1946, *Phys. Rev.*, **70**, 665.
 VAN VLECK and WEISSKOPF, 1945, *Rev. Mod. Phys.*, **17**, 227.

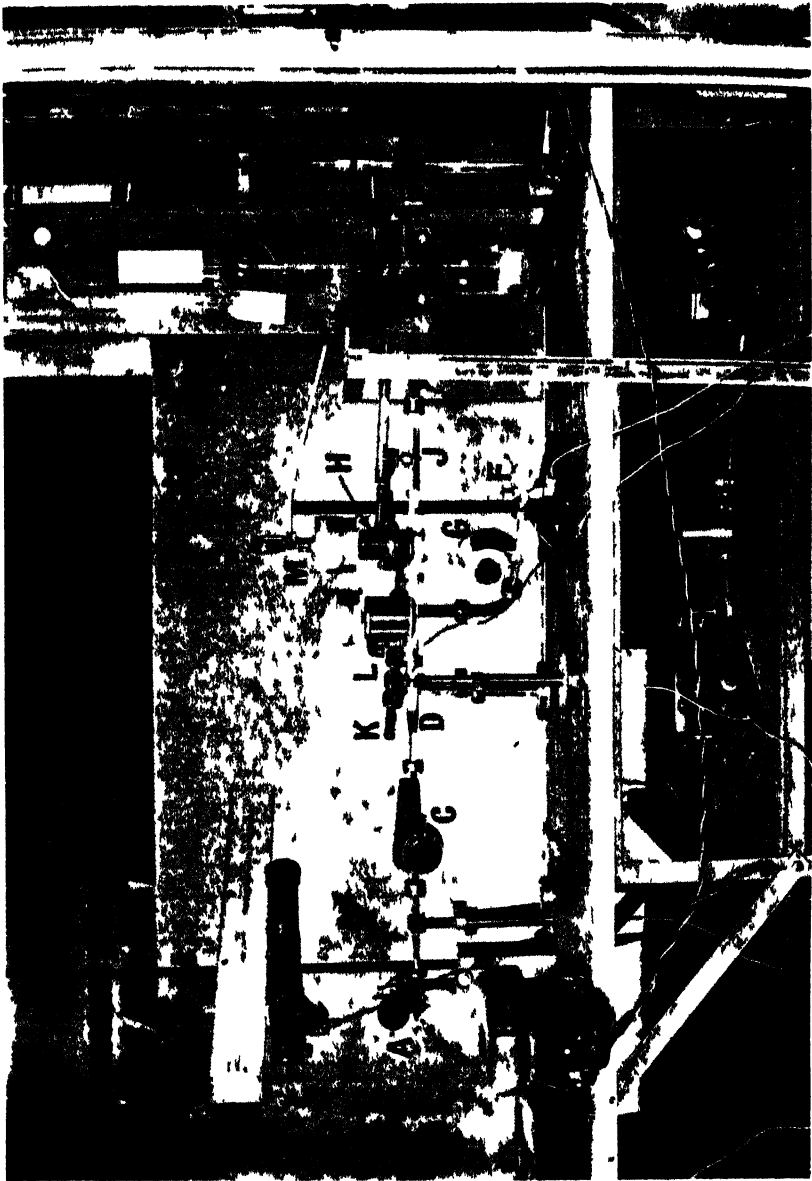


Figure 2 The apparatus

- | | |
|---------------------------------|--|
| A Reflexion klystron oscillator | F Cavity wavemeter |
| B Air blower | F K Crystal rectifiers |
| C G I L Variable attenuators | H Resonant cavity (experimental chamber) |
| D Waveguide twist | V Micrometer head |

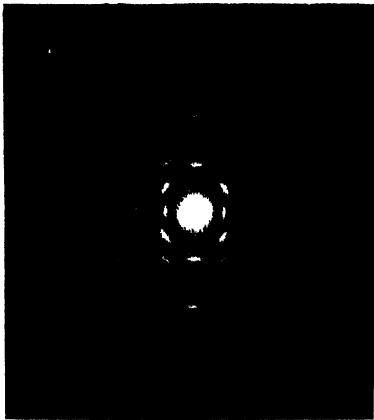


Figure 1 (a) Arc pattern with specimen perpendicular to beam

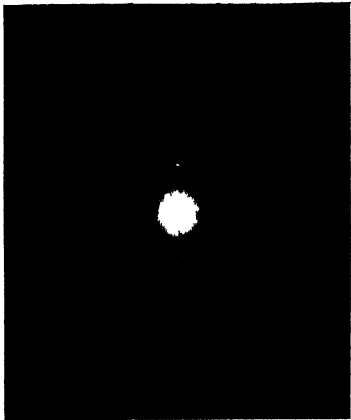


Figure 1 (b) Arc pattern with specimen inclined at 45° to beam

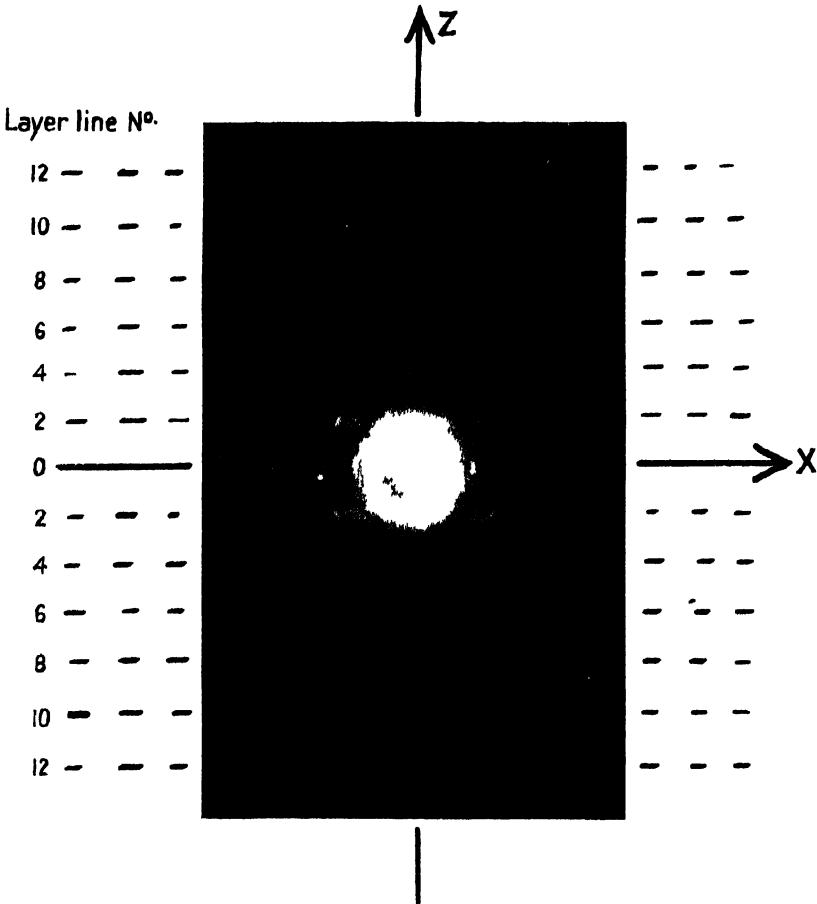


Figure 1 (c) Pattern showing diffuse zones in addition to arcs.

The Molecular Structure and Arrangement in Stretched Natural Rubber

By DOROTHY G. FISHER

Research Fellow of the Research Association of British Rubber Manufacturers at Imperial College

Communicated by G. I. Finch, F.R.S. ; MS. received 30 May 1947

ABSTRACT. Thin films of stretched rubber yield electron diffraction patterns which exhibit, in addition to the Laue spots, a pattern of diffuse zones.

Finch and Wilman (1937) obtained similar composite patterns from various organic crystals and, with Charlesby (1939), they showed that in the case of anthracene the diffuse zone pattern owed its origin to molecules in the crystal, but was similar to that which would be obtained by scattering from independent molecules oriented as in the crystal. The present study has demonstrated that this is also true in the case of rubber, and an appreciation of the significance of the diffuse-zone pattern has led to an advance in our knowledge of the structure of natural rubber.

§ 1. INTRODUCTION

X-RAY fibre diagrams were first obtained from stretched natural rubber by J. R. Katz in 1925. Since then many workers have examined rubber in both the stretched and unstretched state and have made estimates of the lattice constants of the pseudocrystalline structure produced on stretching. C. W. Bunn (1942) recorded an important advance when he arrived at a detailed atomic structure for the extended molecule by appreciating the relative intensity distribution in the x-ray pattern and applying the trial and error method. This structure, though it could not be claimed as final, did not conflict with the observed facts. Other workers, notably Jeffrey, with Cox and Bateman (1942, 1943, 1944), realising that the experimental data available from a fibre diagram are insufficient for a unique Fourier synthesis of the atomic positions, have preferred to examine the crystal structure of relatively simple materials containing isoprene units or related atomic groupings from which the configuration of the elementary units in the poly-isoprene chain might be inferred. It is clear, however, that we are still far from having reached finality in the determination of the molecular structure and arrangement in natural rubber.

Electron diffraction patterns have been obtained from both stretched and unstretched natural rubber (Trillat and Motz 1934, Krailov 1935), but have hitherto added little to the data already provided by x-rays, chiefly owing to the fact that the intensity distribution in the electron diffraction patterns obtained from thin films is so markedly influenced by the relaxation of the third Laue condition.

In 1937 Finch and Wilman first drew attention to the fact that the transmission spot patterns obtained from single crystals of a number of aromatic compounds were associated with a remarkable pattern built up of diffuse areas of scattering,

and they recognised that this diffuse scattering pattern originated from the molecule and was not due to the lattice as a whole. Indeed, in 1939, with Charlesby, they showed that this diffuse-zone pattern was due to molecules which, though oriented with respect to the beam in accordance with the crystal orientation, were apparently executing thermal vibrations about their mean lattice positions. The pattern resembled, except for a slowly varying radial factor, that to be expected if the electron beam had been scattered by independent isolated molecules which had the same orientation as in the normal crystal. It is clear that this diffuse background pattern should, on account of its molecular origin, help in determining both the molecular structure and the molecular orientations in organic crystals which exhibit, as in anthracene, sufficient regularity of atomic arrangement within the molecule. It has now been found possible to obtain well-marked diffuse background patterns from stretched raw rubber, and these have thrown fresh light on the structure and arrangement of the molecules in natural rubber.

§ 2 EXPERIMENTAL

Transmission patterns were obtained from films which were stretched to about ten times their original length, the thickness of the film before stretching being of the order of 2000 Å., as estimated from the interference colours under oblique illumination. The films were prepared by spreading a few drops of a solution of rubber in benzene on a clean water surface in a Langmuir trough; two chromium-plated bars were then placed across the trough and drawn apart, thus stretching the rubber film adhering to their lower surfaces. A portion of the stretched film was then picked up on a grease-free nickel gauze and examined in a Finch-type electron diffraction camera working with a camera length of 47 cm. and an accelerating voltage of 50–60 kv.

The types of pattern obtained are illustrated in figures 1 (a), (b), (c). Generally speaking all the patterns showed evidence of both the normal crystalline and the diffuse-zone patterns. The most favourable exposure conditions for bringing out the two were, however, different, a comparatively long exposure being required for the diffuse background pattern. Thus figures 1 (a) and (b) show the crystalline patterns obtained with a short exposure (about 10 sec.), with the film perpendicular to the beam for figure 1 (a) and inclined at an angle of 45° for figure 1 (b), the fibre axis remaining perpendicular to the beam in both cases, while figure 1 (c) shows a typical diffuse-zone pattern of long exposure (30 sec.) taken with the film perpendicular to the beam. Well-defined diffuse-zone patterns have not yet been obtained with the film sufficiently inclined to the beam to make any appreciable difference to this pattern, since the greater effective thickness of the film gave rise to sundry confusing effects such as a charging up of the specimen and loss of orientation under the prolonged electron bombardment necessitated by this greater thickness. It is also probable that the diffuse background pattern would still not be given prominently by the steeply inclined films, even if these disturbing influences could be eliminated, because in the case of a single crystal the intensity of the background pattern relative to that of the spot pattern should, according to Charlesby, Finch and Wilman (1939), vary inversely with the number of atoms in the crystal flooded by the electron beam, which is in this case proportional to the effective thickness of the film.

§ 3. THE ARC PATTERN

In appreciating the electron diffraction patterns the first step was to ascertain what measure of agreement existed between the arc pattern and the x-ray fibre diagram. A number of unit cells differing in detail have been proposed by various x-ray workers, but as C. W. Bunn (1942) has made the only complete structure analysis, his work was taken as giving the best available data; his proposed unit cell is monoclinic with $a=12.46$ Å., $b=8.89$ Å., c (orientation axis) $=8.10$ Å., $\beta=92^\circ$, and space group $P2_1/a$.

It is immediately clear from the difference between figures 1 (a) and 1 (b) that the crystalline regions in the rubber films examined here have a strongly preferred orientation in the plane of the rubber film. A rough calculation indicated that this orientation was such that the a and c axes tended to lie in the plane of the rubber film. Working on this basis and using, for a first trial, the orthorhombic cell with $a=12.5$ Å., $b=8.9$ Å., $c=8.1$ Å., $\beta=90^\circ$, which is a slightly simplified version of Bunn's proposed cell, it was found possible to index all the spots appearing in the present work (table 1). There were no results specifically requiring the monoclinic cell, but the possibility of this being the true cell cannot be ruled out, as a deviation of 2° in β would not in general be observable in view of the angular spread of the arcs caused by imperfections in the orientation. Bunn himself only introduced this deviation of β from 90° at a late stage in his calculations in order to account for one weak spot which is unlikely to occur in these patterns because the specimen was stationary during the recording of the pattern. The systematic absences, while conforming to the requirements for the space group $P2_1/a$, would equally well satisfy the higher and true orthorhombic symmetry of $P2_12_12_1$ requiring spots of types (h00), (0k0), or (00l) absent for odd values of h , k , or l ; but again, reference to Bunn's calculated intensity values indicates that while (00l) spots are theoretically present for odd values of l they are so weak as to be unlikely to appear in the recorded pattern except in unusually favourable circumstances. The intensity distribution among the arcs in the electron diffraction patterns, figures 1 (a) and 1 (b), cannot be rigorously compared with that of the x-ray pattern, as it is influenced by the extent of the range of orientations present in the specimen, the relaxation of the third Laue condition, and the fact that the specimen was examined in selected stationary positions, thus enhancing the intensity from sets of planes which were preferentially oriented in a direction parallel to the beam. It is nevertheless of interest to note that, making some allowance for these factors, there is good qualitative agreement in intensity distribution between the electron diffraction patterns and the normal (00l) oriented x-ray fibre diagram (table 1). Thus it appears that, while the electron diffraction arc pattern would fit a rather simpler cell having the orthorhombic form and space group $P2_12_12_1$, there is no evidence which conflicts with the small change in β favoured by Bunn and the space group $P2_1/a$; furthermore, the intensity distribution is in qualitative accord with that of the x-ray patterns.

With regard to the absolute value of the identity period, however, a real discrepancy between the electron diffraction results and Bunn's values appears. Thus the best estimates of a and c measured against the graphite standard (110) ring, 1.228 kx. (Nelson and Riley 1945), were 12.46 ± 0.02 kx. and 8.23 ± 0.02 kx., compared with Bunn's values of 12.46 Å. and 8.10 Å. respectively, from which it is seen that the present estimate of the identity period differs significantly from

Table 1. Comparison of observed spacings with those calculated for orthorhombic cell having $a=12.5$ Å., $b=8.9$ Å., $c=8.1$ Å., and comparison of observed intensities with those found by Bunn.

Indices (hkl)	Electron diffraction patterns				X-ray patterns		
	Calculated	Spacings		Intensities observed		Intensities found by Bunn for (001) oriented fibre diagram	
		Observed				Observed	Calculated
		Specimen perpendicular to beam	Specimen at 45° to beam	Specimen perpendicular to beam	Specimen at 45° to beam		
200	6.25	6.21	6.19	VS	—	S	64.5
120	4.19	4.18	4.14	MW	S	VS	254
400	3.12	3.08	—	M	—	M	15
*240	2.10	—	—	—	—	—	1
600	2.08	—	—	—	—	—	0
800	1.56	1.54	—	W	—	VVW	4.5
920	1.33	1.31	—	VW	—	VW	15
1000	1.25	1.23	—	W	—	VVW	7.5
111	5.40	—	5.35	—	VW	VW	2
201	4.95	4.92	4.89	VS	VW	S	24
121	3.72	—	3.69	VVW	VW	S	27
311	3.42	3.39	3.37	VVW	VW	MW	5
401	2.92	—	—	—	—	—	0
521	2.85	—	—	VVW	—	—	0
*231	2.54	—	—	—	—	—	3
611	1.97	—	—	VW	—	MW	0.5
711	1.71	1.70	—	VVW	—	VW	1
801	1.53	1.51	—	VW	—	VW	1
002	4.05	4.05	4.05	VS	VS	MW	8
*012	3.69	—	—	—	—	MW	7
112	3.54	—	3.51	—	VW	MW	3
202	3.40	3.40	—	S	—	MS	26
*022	3.00	—	—	VVW	—	—	6
*122	2.91	—	—	—	—	M	0.5
312	2.76	2.75	2.73	VW	W	M	4
402	2.47	2.48	—	W	—	VVW	1
*322	2.43	—	—	W	—	—	2
412	2.38	—	—	—	—	—	7
132	2.35	2.38	2.36	VW	VW	MW	0.5
512	2.07	2.06	2.04	MS	VW	M	6
432	1.90	—	—	—	—	—	2
242	1.86	—	1.88	—	VW	M	0
*612	1.81	—	—	VW	—	VVW	2
113	2.53	—	2.50	—	VW	—	11
203	2.48	2.51	—	VVW	—	MW	1
*123	2.27	—	—	—	—	MW	6
323	2.02	—	—	—	—	—	—
403	2.04	2.04	—	VW	—	MW	3
*133	1.97	—	—	—	—	MW	1
004	2.02	2.02	2.02	S	VS	MW	9
*124	1.82	—	—	—	—	—	6
314	1.78	—	1.81	—	—	MW	0
205	1.57	1.58	—	VW	—	—	—
506	1.19	—	—	—	—	—	—
516	1.18	1.17	—	VW	—	—	—
008	1.01	1.01	—	VW	—	—	—

* Weak diffractions occurred on photographs from a specimen oriented with its plane at 60° to the beam

† Values given here are rough means of hkl and $\bar{h}\bar{k}\bar{l}$ values.

Bunn's value of 8.10 Å, although the two values for the a axis are in good agreement. These measurements were made on strong arcs of the (h00) and (001) type and the graphite comparison ring was introduced by permitting a drop of diluted graphite suspension in water to spread and dry out on the stretched rubber film. The third axial length b was not investigated in detail as no spots of the (0k0) type appear in transmission patterns on account of the strongly preferred orientation with this axis perpendicular to the plane of the rubber film.

Notwithstanding the difference in the absolute value of the identity period, the general agreement in form between the electron diffraction and x-ray results justifies the use of Bunn's atomic coordinates as a basis for a preliminary investigation of the molecular origin of the diffuse-zone pattern.

§ 4 THE DIFFUSE-ZONE PATTERN

The pattern reproduced in figure 1 (c) shows three distinct features: these are, (i) the haloes characteristic of the amorphous or liquid state, (ii) a pattern of arcs forming a typical crystal fibre diagram, and finally, (iii) a number of diffuse zones, the so-called diffuse background pattern ascribed by Charlesby, Finch and Wilman (1939) to molecular displacement by thermal motions. It is in these zones and their relation to the crystalline pattern that the main interest of this work lies.

Since the diffuse zones in the inner regions of the pattern are associated with clusters of arcs it might at first be thought that they arose from some distortion of the crystalline lattice giving rise to broadening of the arcs. Two considerations, however, indicate that this cannot be the whole explanation of the diffuse-zone pattern; these are that the broadening does not increase with increasing angle of diffraction, and that the diffuse zones occur in the outer regions of the pattern where there is no sign of associated arcs. Consequently we are led to some further consideration of the observed pattern in relation to the characteristics which would be expected on the molecular thermal vibration theory.

In connection with their work on anthracene, Charlesby, Finch and Wilman (1937, 1939) developed a theoretical expression for the intensity arising from a group of rigid molecules oriented in accordance with the crystalline arrangement, but undergoing thermal vibrations. They found that the resulting pattern should consist of two parts: (i) the Laue spot pattern characteristic of the periodicities of the crystal lattice, and (ii) a pattern of diffuse zones characteristic of the atomic arrangement within the molecules, and that the intensities of the diffuse zones should increase relatively to those of the Laue spots with increasing angle of diffraction. These conclusions were borne out by their diffraction patterns from anthracene, and the present rubber patterns show closely similar features, for it has been shown that the arc pattern corresponds to the normal crystalline arrangement, and it is readily apparent from figure 1 (c) that the diffuse zones become increasingly more intense relative to the arcs with increasing angle of diffraction, in fact, they remain prominent in the outer regions of the recorded pattern where the arcs have ceased to be even faintly visible.

Thus the evidence points strongly towards the conclusion that the observed diffuse-zone pattern has this molecular origin. In what follows it is assumed that this is the case and consequently that the arrangement of the diffuse zones is characteristic of the atomic arrangement within the molecules and should

therefore enable us to define the structure of rubber more closely than has hitherto been possible.

In order to use the diffuse-zone pattern in this way it is first necessary to consider the general form of the molecules involved and to make certain broad assumptions as to the manner in which they may be expected to vibrate. It is generally accepted that natural rubber is a long chain polymer of isoprene, C_5H_8 , and that its rubber-like properties arise from the great flexibility which the molecules exhibit on account of the freedom of rotation of consecutive units about single bond linkages. It is therefore obvious that the theory developed for anthracene, in which the molecules were treated as rigid vibrating entities cannot be applied directly to the long flexible molecules of rubber; a fundamental treatment of the problem requires some account to be taken of the relative movements of the atoms within the molecule. As a beginning, however, the simplified case can be examined where the molecule is regarded as being divided into fairly short segments which undergo displacements relative to one another as a result of thermal vibrations while the atoms within a segment remain at rest relative to one another.

The question then arises as to what length of the molecule can be considered to be included in a single segment. One obvious choice would be to regard each isoprene unit as a virtually rigid vibrating entity. There would, however, be some considerable simplification in the work of interpreting the diffuse-zone pattern if it were supposed to arise from a longer length of the molecule, since such a segment, including a number of identical units, would show a greater degree of geometrical regularity. It therefore seems worth considering whether any such choice would be reasonable, and in this connection it is necessary to consider the arrangement of the molecules in an extended specimen of rubber. The stretched material is believed to contain a number of pseudocrystalline regions where sections of the chain molecules, oriented parallel to one another by the stretching, have linked together to form a three dimensional array; these regions are interspersed with disordered regions where the molecules are twisted and tangled in an entirely random manner. The length of a single molecular chain is about 10 000 Å. (Staudinger and Bondy 1930), while the average linear extent of a crystalline region, as indicated by the breadth of the arcs in the x-ray fibre diagram, has been estimated to be of the order of 600 Å. (Hengstenberg and Mark 1928). Thus a single chain is likely to pass through several regions of order and disorder. The effect on the fibre diagram of increasing the extension is to enhance its intensity relative to that of the halo pattern without in any way altering the crystalline spacings; thus it would appear that further extension causes more of the material to take up the ordered arrangement while leaving those regions which were already crystalline unaffected. On the other hand, the disordered regions are never entirely smoothed out however great the extension, as is apparent from the persistence of the halo pattern right up to the maximum extension which the material will maintain without fracture.

It would thus appear that we can legitimately regard the extended rubber specimen as comprising large numbers of comparatively rigid crystalline regions interlaced with disordered regions of greater mobility on which the continued extensibility of the material depends. From this argument it is apparent that there is some justification for supposing that the part of any chain molecule included in a single crystalline region might, to a first approximation, undergo

thermal vibrations as a rigid unit. Such a segment would have a definite repeat unit in the direction of the orientation axis (the identity period of the arc pattern), and could only give rise to diffractions along the same lines as the arcs in the crystal pattern. Reference to the pattern, figure 1 (c), shows that, except for a small angular spread similar to that observed in the arc pattern, the zones do in fact lie near these same lines, thus giving experimental support to the hypothesis. Therefore, in the discussion which follows, the diffuse-zone pattern has been treated as though it arose from molecular segments containing at least 10 to 15 repeat units and with a total length of some 100 Å., each behaving as a self-contained array of scattering centres diffracting quite independently of any adjacent chain or any other segment in the same chain. It is nevertheless appreciated that once the general lay-out of the molecule has been defined some further refinements may be obtained by taking into account the relative movements of atom groups within a segment and considering the pattern to be anticipated from individual isoprene units. An investigation of this kind is now being undertaken.

The interpretation of the pattern has been approached by the trial and error method, and in order to do this it was first necessary to make some deductions, based on reasoning from other evidence, as to the atomic structure of the molecule. As mentioned above, rubber is a long chain polymer of isoprene, C_5H_8 ; the identity period along the extended chain is such that it must include two isoprene units, and Bunn (1942) has shown conclusively that it takes up the *cis* form in relation to the double bonds. By analogy with the long chain hydrocarbons it might be anticipated that the chain would be planar, but the identity period along the axis of elongation, which all observers agree to be in the region of 8.1–8.2 Å., is too short to permit this without serious and improbable distortions of the accepted bond lengths and angles; consequently we are forced to the conclusion that the molecule is non-planar. There are, then, many ways in which the chain can be supposed to be kinked and twisted to give the observed identity period. For the present work, however, two models have been selected for investigation: the first is the structure proposed by Bunn which was taken because it was the only complete structure so far proposed on the basis of direct experimental results; the second, hypothetical, structure has been chosen as involving the minimum of distortion from the wholly planar form, and therefore forming a useful basic model on which further modifications can be imposed if necessary. It is a model in which all the carbon atoms of each isoprene unit are planar and the planes of consecutive units are parallel to each other and to the orientation axis; furthermore, all bond lengths and angles are normal, the reduction in identity period being attained by straining the connecting links between consecutive units out of the plane, so that the molecule forms a series of steps up and down as indicated diagrammatically in figure 2. This was, in fact, the molecule selected by Bunn as the initial basis for his calculations, but in order to satisfy his x-ray spot intensities he subsequently modified it considerably by introducing certain distortions in the bond lengths and angles and moving the methyl group out of the main plane of the isoprene unit. The atomic coordinates for one complete repeat unit of each structure are shown in tables 2 and 3, and their projections along three rectangular axes in figures 3 and 4.

The problem then is to find the positions of the intensity maxima arising from a chain segment comprising ten to fifteen such units. In computing such a pattern we have to take into account two factors: the form of the individual

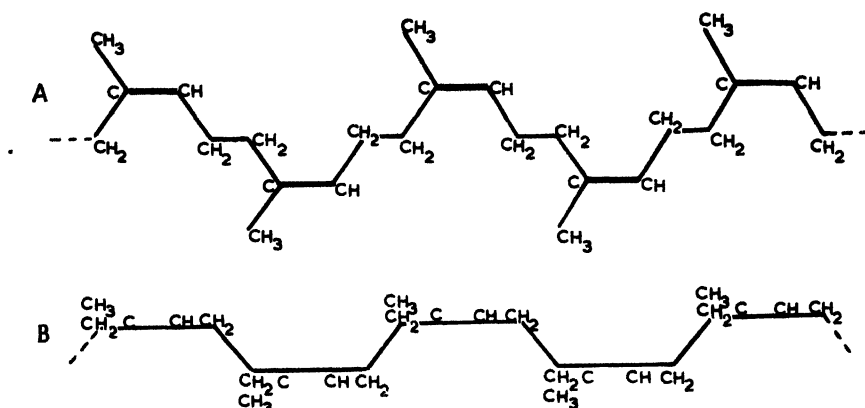


Figure 2. Rubber molecule—diagrammatic.

A. Projection in plane of double bonds.

B. Projection perpendicular to double bonds.

Table 2. Atomic coordinates (A), for Bunn's molecule—Single repeat unit

(a) As given by Bunn, *Proc. Roy. Soc. A*, 1942, **180**, 56. Left-handed "up" molecule. Referred to crystallographic axes $a = 12.46$, $b = 8.89$, $c = 8.10$ Å., $\alpha = \gamma = 90^\circ$, $\beta = 92^\circ$.

	C_1 (CH ₂)	C_2 (C)	C_3 (CH)	C_4 (CH ₂)	C_5 (CH ₂)	C_6 (CH ₂)	C_7 (C)	C_8 (CH)	C_9 (CH ₂)	C_{10} (CH ₂)
x_1	0.753 <i>a</i>	0.854 <i>a</i>	0.845 <i>a</i>	0.745 <i>a</i>	0.958 <i>a</i>	0.744 <i>a</i>	0.644 <i>a</i>	0.659 <i>a</i>	0.757 <i>a</i>	0.532 <i>a</i>
y_1	0.899 <i>b</i>	0.865 <i>b</i>	0.905 <i>b</i>	0.959 <i>b</i>	0.876 <i>b</i>	0.834 <i>b</i>	0.874 <i>b</i>	0.905 <i>b</i>	0.834 <i>b</i>	0.828 <i>b</i>
z_1	0.802 <i>c</i>	0.703 <i>c</i>	0.542 <i>c</i>	0.457 <i>c</i>	0.773 <i>c</i>	0.326 <i>c</i>	0.215 <i>c</i>	0.052 <i>c</i>	0.025 <i>c</i>	0.268 <i>c</i>

(b) Referred to three orthogonal axes x , y , z , and to C_1 as origin. y and z are parallel to b and c respectively.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
x	0.000	1.259	1.147	0.100	2.680	-0.112	1.357	-1.172	0.050	-2.755
y	0.000	0.303	0.052	0.533	0.205	0.577	0.222	0.053	0.577	-0.632
z	0.000	0.846	2.145	-2.792	0.328	3.851	-4.703	-6.034	6.702	-4.224

Table 3. Atomic coordinates (A), for simple molecule having plane units and standard bond lengths and angles—Single repeat unit

(a) Referred to orthogonal axes x , y , z , and to C_1 as origin.

	C_1 (CH ₂)	C_2 (C)	C_3 (CH)	C_4 (CH ₂)	C_5 (CH ₂)	C_6 (CH ₂)	C_7 (C)	C_8 (CH)	C_9 (CH ₂)	C_{10} (CH ₂)
x	0	1.26	1.26	0	2.51	0.04	1.22	1.22	0.04	-2.48
y	0	0	0	0	0	1.22	1.22	1.22	-1.22	1.22
z	0	0.89	-2.22	-3.11	0	-4.05	-4.94	6.27	-7.16	4.05

(b) Referred to axes x , y , z as above, and to the midpoint of $C_4 - C_6$ as origin.

	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_{10}
x	-0.02	1.24	1.24	-0.02	2.49	0.02	-1.24	-1.24	0.02	-2.50
y	0.61	0.61	0.61	0.61	0.61	-0.61	-0.61	-0.61	-0.61	-0.61
z	3.58	2.69	1.36	0.47	0.58	-0.47	-1.36	-2.69	-3.58	-0.47

unit, and the existence of a number of such units along the segment forming effectively a line diffraction grating with a definite identity period. It is only at those positions where both conditions are favourable that maxima can occur. The line grating alone would give rise to maxima along regularly spaced lines perpendicular to its own length, while the individual unit would give a more

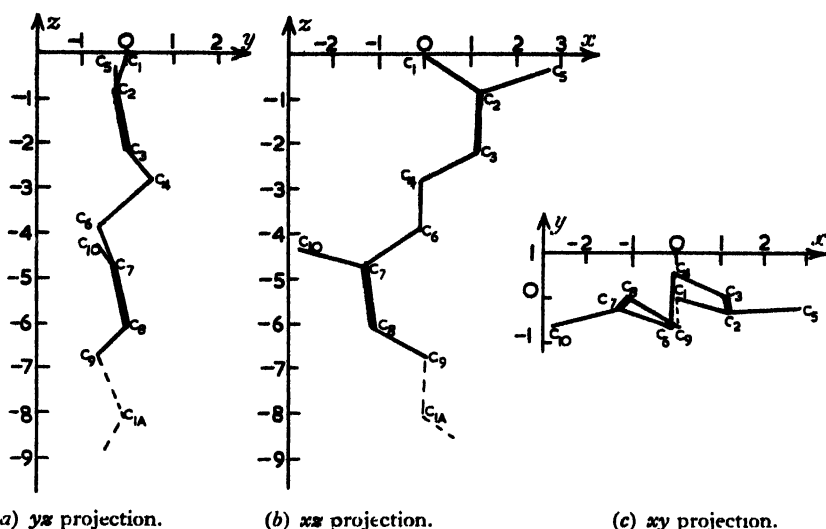


Figure 3. Rubber molecule proposed by Bunn. Projections along three orthogonal axes of one left-handed "up" molecule (Bunn's nomenclature); z axis in direction of extension.

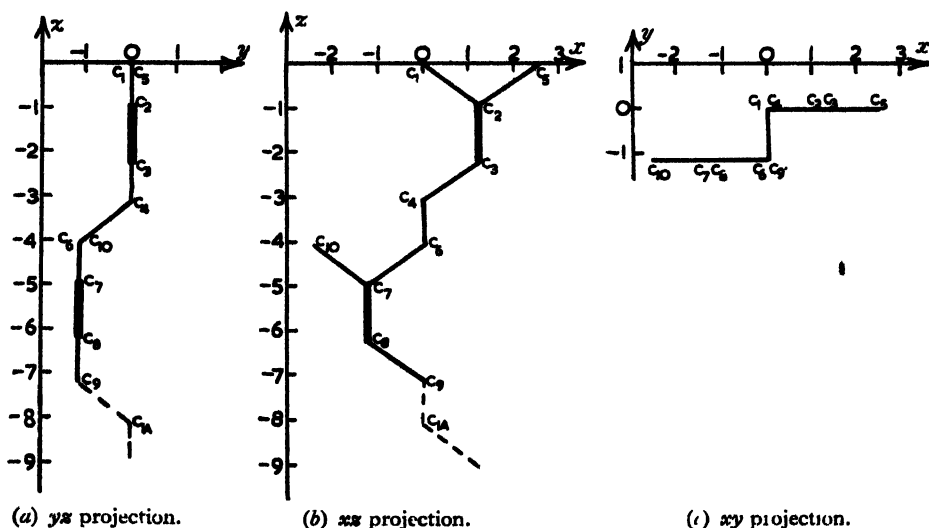


Figure 4. Suggested rubber molecule with planar isoprene units and standard bond lengths and angles. Projections along three orthogonal axes: z axis in direction of extension.

complex two-dimensional pattern. Consequently, in the computations which follow, the method adopted has been to select the lines along which the identity period permits maxima to occur and to examine the intensity variations which the atomic arrangement in the single unit gives rise to along these lines. In other words, the calculations are made as for a single unit, but the repetition

of this unit along the molecular segment introduces a geometrical regularity which limits the regions over which the resulting pattern need be considered.

Since the molecular segment is non-planar, it is no longer possible to construct a single two-dimensional intensity contour map adaptable to any crystal orientation by simple linear transformation of the axial lengths as was done in the case of anthracene; instead, each orientation requires a separate contour diagram. However, the only diffuse-zone patterns considered in this work were taken with the plane of the rubber film effectively perpendicular to the electron beam, and the arc patterns, figures 1 (a) and 1 (b), indicate a strongly preferred orientation with the ac crystallographic plane lying in this plane. Consequently we are principally concerned with this one orientation for which the x and z axes of atomic coordinates (which are in the ac plane), figures 3 and 4, are perpendicular to the electron beam and the y axis is parallel to it. In computing the anticipated intensity distribution for this orientation it is only necessary, as will be shown below, to take into account the x and z atomic coordinates of the molecule. More generally, for other orientations, the diffraction pattern can be considered as practically that which would be obtained from the projection of the molecule in the plane perpendicular to the electron beam.

The (010) orientation in the film plane is not quite perfect; thus the appearance of certain spots of the general (hkl) type in the pattern taken with the film perpendicular to the beam (table 1) indicates a partial rotation about the c axis. In addition, the c axis itself shows some angular deviation from its mean position within the ac plane as is indicated by the drawing out of all spots into arcs of about 10° angular range, and it seems likely that it will similarly suffer some angular deviation in the plane perpendicular to this. The most probable situation is something approaching a Gaussian distribution of orientation with its maximum frequency in the (010) orientation, and such a state of affairs would be expected to result in some broadening of the intensity maxima, but should not give rise to any peaks not anticipated for the preferred orientation. If the angular spread of direction of the b axis is similar to that exhibited by the c axis in the (010) plane the broadening will not be appreciable since an angular deviation of $\pm 5^\circ$ about an axis normal to the electron beam produces little change in the projection of the molecule normal to the beam. In what follows, therefore, the pattern has been computed for the single orientation with the xz plane normal to the electron beam.

The calculation of intensities is based on the following argument (see figure 5):—

Suppose the amplitude of the wave diffracted by any atom p in a direction S'

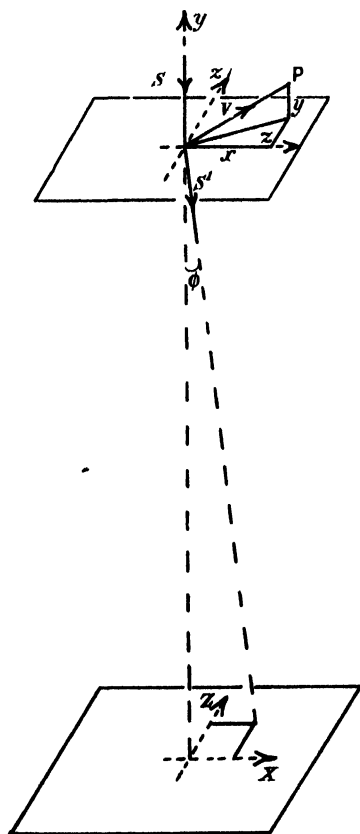


Figure 5. Illustration for derivation of intensity relationship.

from a wave incident in a direction \mathbf{S} to be E_p (where \mathbf{S} and \mathbf{S}' are unit vectors). The intensity of the wave resulting from a group of such atoms is obtained by adding the amplitudes of the constituent diffracted waves, having due regard to their phases, and squaring the result.

Thus the resultant intensity is given by the expression

$$\{\sum_p E_p \cos (2\pi d_p/\lambda)\}^2 + \{\sum_p E_p \sin (2\pi d_p/\lambda)\}^2,$$

where d_p is the path difference between the wavelet scattered from p and from an arbitrary origin, and λ is the wavelength of the radiation considered.

For an atom whose position relative to the origin is given by the vector \mathbf{v}_p we have

$$\begin{aligned} d_p &= \mathbf{v}_p \cdot (\mathbf{S}' - \mathbf{S}) \\ &= (\mathbf{x}_p + \mathbf{y}_p + \mathbf{z}_p) \cdot (\mathbf{S}' - \mathbf{S}) \\ &\approx (\mathbf{x}_p + \mathbf{z}_p) \cdot (\mathbf{S}' - \mathbf{S}) \text{ since } \mathbf{y}_p \cdot (\mathbf{S}' - \mathbf{S}) \approx 0 \end{aligned}$$

for small ϕ .

If the directions \mathbf{S}' and \mathbf{S} strike the photographic plate in points X , Z and O , referred to axes in the plate we have

$$\begin{aligned} \mathbf{S}' &= (\mathbf{X} + \mathbf{Z} + \mathbf{L}) / \sqrt{X^2 + Z^2 + L^2} \text{ where } L \text{ is the camera length} \\ &\approx (\mathbf{X} + \mathbf{Z} + \mathbf{L}) / L \quad \text{since } \phi \ll 2^\circ \end{aligned}$$

and

$$\mathbf{S} = \mathbf{L} / L$$

so that

$$(\mathbf{S}' - \mathbf{S}) = (\mathbf{X} + \mathbf{Z}) / L.$$

Hence, the expression for the intensity becomes

$$\begin{aligned} &\{\sum_p E_p \cos [(2\pi/\lambda) (\mathbf{x}_p + \mathbf{z}_p) \cdot (\mathbf{X} + \mathbf{Z}) / L]\}^2 + \{\sum_p E_p \sin [(2\pi/\lambda) (\mathbf{x}_p + \mathbf{z}_p) \cdot (\mathbf{X} + \mathbf{Z}) / L]\}^2 \\ &= \{\sum_p E_p \cos [(2\pi/\lambda L) (x_p X + z_p Z)]\}^2 + \{\sum_p E_p \sin [(2\pi/\lambda L) (x_p X + z_p Z)]\}^2 \dots \dots (1) \end{aligned}$$

when \mathbf{X} and \mathbf{Z} are taken parallel respectively to \mathbf{x} and \mathbf{z} , and the axes are rectangular.

The amplitude E_p of the wave diffracted from an individual atom depends on the scattering power of the atom and on the angle of scattering, the latter relationship resulting in a falling off in intensity from the centre towards the outer parts of the pattern. With regard to the scattering power of the atoms concerned, the hydrogen atoms have been neglected as separate scattering entities, but have been regarded as adding a little to the effective weight of the carbon atom to which they are attached; thus the scattering factors of the C, CH, CH₂ and CH₃ groups have been taken in the ratio 6:7:8:9, which assumption was also made by Bunn in determining his atomic coordinates.

In computing intensities the following standard conditions have been taken: the factor λL has been taken as 2.75 Å.cm. (for direct comparison with an experimental pattern the dimensions of the computed pattern have been reduced in the ratio $L(\text{observed})/2.75$), and the origin of coordinates has been taken as coinciding with the carbon atom C₁ (figures 3 and 4). Atomic coordinates have been measured in Ångström units, and coordinates on the photographic plate in centimetres.

Thus, substituting for the various known or assumed values, the expression used in the computations of intensity at any point X, Z of the photographic plate becomes, for the Bunn molecule,

$$\left\{ \begin{array}{l} 8 \cos [(360/2.75)(0 - 0)] \\ - 6 \cos [(360/2.75)(1.259 X - 0.846 Z)] \\ - 7 \cos [(360/2.75)(1.147 X - 2.145 Z)] \\ - 8 \cos [(360/2.75)(-0.100 X - 2.792 Z)] \\ - 9 \cos [(360/2.75)(2.680 X - 0.328 Z)] \\ - 8 \cos [(360/2.75)(-0.112 X - 3.851 Z)] \\ - 6 \cos [(360/2.75)(-1.357 X - 4.703 Z)] \\ - 7 \cos [(360/2.75)(-1.172 X - 6.034 Z)] \\ - 8 \cos [(360/2.75)(0.050 X - 6.702 Z)] \\ - 9 \cos [(360/2.75)(-2.755 X - 4.224 Z)] \end{array} \right\}^2$$

$$+ \left\{ \begin{array}{l} 8 \sin [(360/2.75)(0 - 0)] \\ - 6 \sin [(360/2.75)(1.259 X - 0.846 Z)] \\ - 7 \sin [(360/2.75)(1.147 X - 2.145 Z)] \\ - 8 \sin [(360/2.75)(-0.100 X - 2.792 Z)] \\ - 9 \sin [(360/2.75)(2.680 X - 0.328 Z)] \\ - 8 \sin [(360/2.75)(-0.112 X - 3.851 Z)] \\ - 6 \sin [(360/2.75)(-1.357 X - 4.703 Z)] \\ - 7 \sin [(360/2.75)(-1.172 X - 6.034 Z)] \\ - 8 \sin [(360/2.75)(0.050 X - 6.702 Z)] \\ - 9 \sin [(360/2.75)(-2.755 X - 4.224 Z)] \end{array} \right\}^2; \quad \dots\dots(2)$$

and for the second, less distorted, molecule, a similar expression holds but with the appropriate atomic coordinates substituted for the Bunn coordinates.

With an identity period of 8.1 Å. and with $\lambda I = 2.75$ the "layer-line" separation is $2.75/8.1$ cm.; in other words the lines near which our preliminary assumptions permit intensity maxima to occur have Z values (taken to the nearest $\frac{1}{2}$ mm.) of 0, 0.35, 0.7, 1.05, 1.35, 1.7, 2.05, 2.4, 2.75 etc. cm. Inspection of the photographs, however, suggests that the most pronounced maxima lie on the even-order layer lines (as might be expected from the similarity in form of the two halves of the molecule) and consequently the computations have been confined to those lines. They were carried out in general at 2 mm. intervals along these lines except in a few cases where intermediate values were obtained. Angles in the structure-factor expression were taken to the nearest degree.

The molecule has four possible orientations in the film, for the conditions governing its orientation are that the c axis direction (z axis of coordinates) shall lie in the direction of extension, and that the plane of the isoprene units (xz projection) shall be parallel to the plane of the film. From figures 3 (*a*) and 4 (*a*), which show the appropriate projection, these conditions are seen still to be fulfilled if the whole molecule is rotated through 180° about the x or z axis or both. Taking the axes of coordinates as fixed in relation to the film this is equivalent to changing the sign of either the x or z coordinates, or both. Now it is reasonable to suppose that any section of the film impinged on by the electron beam will include equal numbers of molecules in all the possible orientations; consequently in computing it is necessary to consider the intensities arising from each of these orientations, the resulting intensity at any point being the sum of that due to the separate constituents taking into account the basic assumption

of no phase relationship between adjacent molecules. In fact it is only necessary to consider two of the orientations because, owing to the squares, the expression (1) for the intensity arising from a single molecular unit is not affected when the signs of both the atomic coordinates are reversed. It is also only necessary to consider one quadrant of the pattern because the effect on the expression (1) of reversing the sign of either X or Z is precisely the same as that of reversing the sign of the corresponding atomic coordinates. In the case of the undistorted molecule the atomic configuration shows some degree of symmetry which is best seen from table 2 (*b*) in which the origin of coordinates has been moved to the midpoint of the repeat unit. This introduces a further simplification to the computations, for the two orientations x, z and $x, -z$ now give identical intensity distributions as long as we confine ourselves to the layer lines; for, provided that the ten atoms of the repeat unit are regarded as repeating indefinitely (which is what confining our attention to the layer lines implies), a change from the x, z to the $x, -z$ orientation merely implies a change of origin from the point C_1 to the point C_8 (table 3), which cannot have any effect on the computed pattern since the choice of origin is purely arbitrary. Therefore for the undistorted model calculations were carried out for the x, z orientation only, and the values were multiplied by two to compare with the summed intensities for the two orientations of the Bunn molecule.

§ 5. COMPARISON OF COMPUTED PATTERNS WITH OBSERVED PATTERN

The computed intensity distributions along the zero, 2nd, 6th, 8th, 10th and 12th layer lines are shown in figures 6 (*a*)–(*f*). The fourth layer line has been omitted as it showed no pronounced maxima. An attempt has been made to construct the patterns which would be anticipated from the positions of these maxima, and the results are shown in figures 7 (*a*) and 7 (*b*), together with a trace taken from an experimental pattern, figure 7 (*c*). In constructing these patterns it has been assumed that the visible breadth of the diffraction maxima corresponds to the half-maximum breadth of the intensity peak, and that the imperfections of molecular orientation are of the order indicated by the extent of the arcs in the Laue spot pattern, so that the whole pattern suffers an angular spread of about 9° . The length of the molecular segment considered has been taken as sufficient to limit the layer lines to true lines, the apparent breadth of the diffuse zones in the c axis direction arising solely from the small rotation referred to above.

The similarity in form between the computed and experimental patterns is readily apparent and gives ample support to the fundamental theory on which the calculations are based; that is, we are justified in assuming that the observed diffuse-zone pattern arises from molecules which are oriented much as in the crystal but behave as entirely independent groups of scattering centres. In other words, the pattern is characteristic of what would be expected from a group of oriented molecules in a pseudo-gaseous phase.

A close inspection of the curves (figures 6 (*a*)–(*f*)) reveals significant differences between the intensity distributions from the two models. The most striking difference is along the eighth layer line where the two curves of figure 6 (*d*) show entirely different characteristics, the Bunn model giving rise to three weak peaks at roughly 0.5, 1.5 and 2.5 cm. from the centre, while the undistorted model

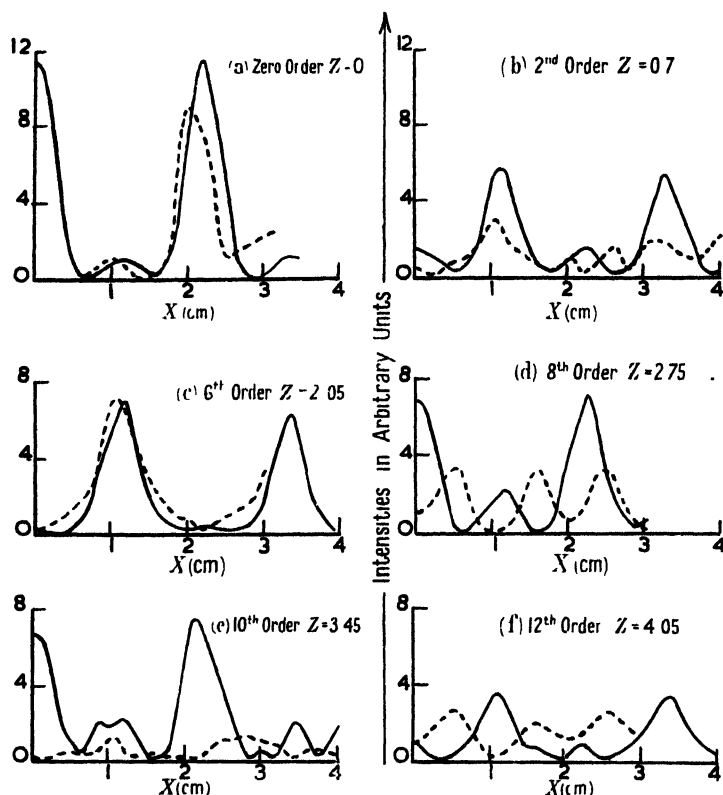


Figure 6. Computed intensities along layer lines for the two atomic configurations investigated

— — — Model with planar units and standard bond lengths and angles
 - - - - - Model due to C. W. Bunn.

gives a central strong peak, and a second of about the same intensity at 2.2 cm. Comparison with figure 1 (c), and with the traced experimental pattern figure 7 (c), shows that the undistorted model is in far better agreement with experiment in this region; in fact, along this line the agreement is exact within the technical limitations of the patterns. The next characteristic difference

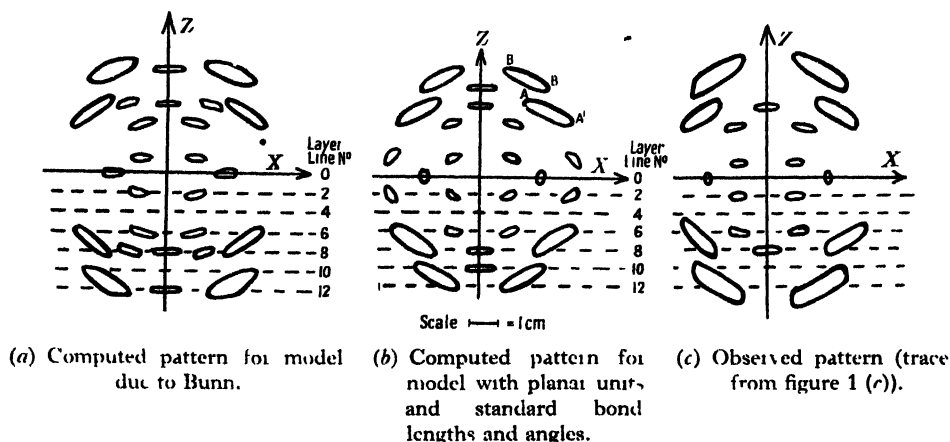


Figure 7. Computed and observed diffuse zone patterns.

between the two models is that the undistorted form gives sharper peaks; thus on the tenth and twelfth layer lines the Bunn model shows no pronounced peaks, but only certain wide zones of slightly increased intensity, while the undistorted model and the experimental pattern both show well defined peaks along these lines. The experimental data and undistorted model pattern are again in fairly good agreement in relation to peaks at 2.2 and 1.1 cm. on the tenth and twelfth orders respectively; these could well coalesce and give rise to the extended zone BB' (figure 7 (b)), though the position of this observed zone is not quite what would be expected from the computed intensities. The error is probably not serious since the observed pattern is very faint and of poor definition in these outer regions, and the computed intensity maxima here are highly sensitive to small changes in the atomic coordinates owing to the large values of their coefficients. Along the zero and sixth order layer lines the two computed patterns are in good agreement with one another and with observation. On the second order line they differ again slightly, the undistorted model still giving more definite peaks though, while the peak at 1 cm. accords well with observation, the moderate peak to be expected at 3.2 cm. is absent from the experimental pattern.

It would thus appear that while both the molecular models investigated give rise to patterns agreeing in their broad general characteristics with those actually observed in the diffuse-zone electron-diffraction pattern, a more detailed examination shows that the computed intensity curves from the undistorted model are in much closer agreement with the facts than is the case with Bunn's model. This agreement is in fact remarkably good, the only discrepancy being the absence from the experimental pattern of a peak to be expected, in the case of the undistorted model, on the second order layer line at 3.2 cm. In particular, the undistorted model gives much the better agreement along the eighth order layer line, where the most sharply defined zones in the experimental pattern occur, and where Bunn's model gives an intensity distribution which is entirely at variance with observation.

§ 6 CONCLUSIONS

Although there are still some minor discrepancies to be resolved, the appearance of the diffuse-zone electron-diffraction pattern arising from thin films of stretched natural rubber is adequately explained on the basis of the thermal oscillations of the long-chain rubber molecules in the crystalline regions, this pattern being analogous with that which would be given by a stream of oriented molecules in a pseudo-gaseous condition. As Charlesby, Finch and Wilman (1939) pointed out the diffuse-zone pattern has distinct advantages in the examination of complex molecules, since it gives a direct indication of the configuration of the molecule independent of its mode of fitting into any particular crystalline lattice. In spite of the diffuseness of the patterns with which we are dealing a study of the diffuse zone or molecular pattern has shown that it is possible to discriminate decisively between two postulated atomic configurations which do not differ greatly.

In the case of natural rubber, neither of the two configurations investigated agrees in every respect with experiment, but the evidence brought forward above supports the view that the atomic arrangement in the stretched rubber molecule approximates to a simple form having planar units and standard bond lengths and angles, rather than to the more complex form hitherto considered.

ACKNOWLEDGMENTS

I wish to thank the Council of the Research Association of British Rubber Manufacturers for the grant of a R.A.B.R.M. Research Fellowship, which enabled this work to be carried out. I am also indebted to Professor G. I. Finch, F.R.S., for his valuable advice and encouragement, and to Dr. H. Wilman for many helpful discussions.

REFERENCES

- BATTMAN, L. and JEFFREY, G. A., 1943, *Nature, Lond.*, **152**, 446.
 BUNN, C. W., 1942, *Proc. Roy. Soc., A*, **180**, 40.
 CHARLESBY, A., FINCH, G. I. and WILMAN, H., 1939, *Proc. Phys. Soc.*, **51**, 479.
 COX, E. G. and JEFFREY, G. A., 1942, *Trans. Faraday Soc.*, **38**, 241.
 FINCH, G. I., 1941, *Proc. Roy. Soc., A*, **179**, 67.
 FINCH, G. I. and WILMAN, H., 1937, *Ergebn. exakt. Naturw.*, **16**, 353.
 HENGSTENBERG, J. and MARL, H., 1928, *Z. Kristallogr.*, **69**, 271.
 JEFFREY, G. A., 1944, *Trans. Faraday Soc.*, **40**, 517.
 KATZ, J. R., 1925, *Chem. Ztg.*, **49**, 353; 1927, *Chem. Ztg.*, **51**, 381.
 KRUILOV, K. I., 1935, *J. Exp. Theor. Phys. (U.S.S.R.)*, **5**, 524.
 NELSON, J. B., and RILLY, D. P., 1945, *Proc. Phys. Soc.*, **57**, 477.
 STAUDINGER, H. and BONDY, H. F., 1930, *Ber. dtsh. chem. Ges.*, **63** (B), 724.
 TRILLAT, J. J. and MOTZ, H., 1934, *C.R. Acad. Sci., Paris*, **198**, 2147.

Reviews of Books

Electronics and their Application in Industry and Research, edited by BERNARD LOVELL. Pp. xvi+660. (London: The Pilot Press Ltd., 1947.) 42s. net.

It is becoming trite to say that during the war the science of electronics developed at a phenomenal pace. Apparatus which, in 1939, was found only in the communications research laboratory, developed almost out of recognition to lead to the growth of an extensive industry concerned with the production of electromagnetic waves in the centimetre band, with the use of infra-red waves, and with the electronic control and measurement of industrial and medical processes. Here, indeed, is an example of the enormous rewards which accompany the devotion of the energy of great teams of scientists, and the coordination of work between Universities, Government laboratories, and technical industry to the theory, design and production of apparatus based on principles which previously were thoroughly understood and used only in the laboratory and by comparatively few workers. It is, however, remarkable that nearly all these advances have not been fundamental in nature, but technical in that they have been concerned with the practical application of electronic principles which had already been enunciated. It is those pre-war pioneers, mostly working in University laboratories, who gradually introduced order and logic to the understanding of the extraordinary versatility of the free electron in a gas at low pressure, to whom this remarkably interesting and informative book should be dedicated.

In a generous allowance of 660 pages there is described here, for the first time in a technical book, an advanced account of those aspects of electronics which developed apace during the war. Written by fourteen experts, the book provides a useful introduction to the theory and practice of the most recent developments in electronic apparatus. Each section is accompanied by an excellent bibliography, detailing the most important relevant papers so that the student can follow up the introduction to that aspect of modern electronics in which he is interested, and gain a reasonably complete knowledge of a particular field.

After an interesting introduction by the editor, Dr. Bernard Lovell, there follows an account of electron physics written by Dr. F. A. Vick. Though evincing that the author has a complete grasp of his subject, yet this section is inevitably condensed to such an extent that

the reader, unless he has already studied the subject in considerable detail, would fail to follow the arguments. It also seems unfortunate that this chapter is somewhat divorced from the rest of the book in that, once the basic electron emission equations have been established, they are not referred to again by any of the other authors: the standard of knowledge of modern physics demanded to appreciate Dr. Vick's contribution is much more advanced than that needed in any subsequent chapter. Indeed, a major criticism of this book is that there has been too little collaboration between the various authors so that there is no development of the subject as a whole as the book progresses. In view of the wide and varied nature of the complete text, this is to be expected to some extent, but one feels that closer links could have been established between the various sections, adding greatly to the book's value to the advanced student in giving a consistent picture of the development of modern electronic devices.

Photo-cells and television pick-up tubes are covered in two sections by authors from the research laboratories of Electric & Musical Industries Ltd., where most of the development work on iconoscopes and orthicons has taken place in this country. Drs. Lubszynski and McGee give competent accounts of their respective subjects, the section on television transmitting tubes being particularly valuable in that, for the first time in a technical book, the principles and practice involved have been properly described.

Dr. A. Elliott has contributed a fascinating chapter on photo-cells used in the infra-red which should prove of great value to those research workers who wish to use the highly sensitive devices developed during the war to establish secret signalling methods and to detect hot body radiation, and harness them for the study of the infra-red as used in spectroscopic analysis, astronomy, hygrometry and gas analysis.

The modern multi-segment magnetron, and the klystron tubes are described by Dr. F. C. Thompson in a chapter "Thermionic Valves for Very High Frequencies". It seems a pity that the excellent account given of the action of these tubes could not be accompanied by extra matter dealing with their use in u.h.f. circuits, particularly since there is such scanty literature available on this subject. The same lack of circuitry is evident in a long section of the book devoted to Radar, written by Dr. R. A. Smith, though this omission is more than compensated by the interest which Dr. Smith stimulates in describing apparatus which, originally used for the investigation of the ionosphere, ultimately led to that extraordinary "box" which enables an observer in a ship or aircraft to obtain a fairly detailed picture on a cathode-ray tube screen of territory hidden by darkness or cloud.

The use of cold cathode valves for the control of industrial apparatus of certain types is described by Mr. I. Atkinson of Ferranti Ltd. Mr. H. Wood, from the same firm, contributes an up-to-date and valuable account of high frequency heating, and also describes an electronic apparatus for measuring moisture content.

Those highly ingenious devices which became known during the war as servo-mechanisms are the subject of a 50-page article by Mr. F. H. Belsey of Metropolitan-Vickers Ltd. Though difficult reading, this chapter is perhaps the best written section of the book; it deals with a complex subject in a logical and mathematical manner leading to a description of the types of gear employed.

In medical science, the applications of electronics are becoming increasingly numerous. Two chapters, "Electronics in Medicine", and "Electronics in Physiology", are contributed by Drs. Grimmett and Pumphrey respectively. The betatron, source of high-energy particles, is the subject of a special section by Dr. J. D. Craggs, who gives a theoretical and practical account of its development and use which clears up many misunderstandings and difficulties.

Finally, Dr. V. E. Cosslett, who is well-known for his research work and book on electron-optics, has written an excellent 100 page description of electron microscopy and electron diffraction.

This book, despite its size and the range of principles and techniques it describes, is singularly free from error; a tribute to the erudition of the various authors, to the editor, and to the excellence of the publisher's work. Unfortunately, as in nearly all technical books, the reader is occasionally obliged in one or two chapters to read the ill-constructed sentences of men who have devoted their lives to becoming technical experts. More than 400 excellent diagrams and photographs illustrate the text. It is a volume that every individual seriously interested in electronics should have on his bookshelf.

J. YARWOOD

New Developments in Ferromagnetic Materials, by J. I. SNOEK. Pp. viii + 136. (Amsterdam: Elsevier, 1947. Distributors for British Empire, Cleaver-Hume Press, London.) 13s. 6d.

The physical and mental frustration during five years of foreign occupation, with an "ever-growing burden of oppression and starvation" might well have dulled the urge to scientific enquiry in the Netherlands. That it did not do so is abundantly clear from this monograph, one of an extensive series on the progress of research in Holland during the war. The work of the author, a distinguished member of the Philips research group, on magnetic materials and much besides, was widely known and appreciated before the war; as was also his genial personality, which is reflected both in the general outlook conveyed by this book, and in the incidental and unobtrusive humour of word or phrase. The work here described is primarily that of the author and his associates at the Philips Laboratories between 1940 and 1945, but in the three main sections of the book the general background is sketched in sufficiently to indicate the setting of the various special investigations.

The first section, dealing with a number of themes under the general heading of the statics of ferromagnetism, includes an account of an experimental survey, by a comparatively rapid approximate method, of the crystal anisotropy and magnetostriction of binary and ternary alloys. A high initial permeability is associated with low values of these quantities, but even in those binary alloys for which zero values of each occur in a single phase range the zeros are unlikely to coincide at a particular composition. For a ternary alloy zero points become zero lines, and even a rough determination of the course of these lines shows whether unusually high permeability characteristics are likely for any composition. In the Fe-Si-Al system, for example, an intersection is indicated in the region where a maximum permeability had been found earlier by purely empirical methods. Although in the systems so far studied such intersections are infrequent, the investigation, which is still proceeding, has provided much valuable information.

Under the dynamics of ferromagnetism there is a discussion of eddy current effects, and of various after effects, tentatively classified as thermic, ionic and electronic. One long standing puzzle has been solved in that an entirely satisfying explanation, verified by experiment, is given of the often reported observations apparently showing that the initial magnetization curve crosses over the ascending branch of the hysteresis curve. The effect is a spurious one which may arise jointly from eddy current and demagnetization effects (it does not occur for ring specimens) to an extent depending on the rapidity with which the field is changed and the magnitude of the steps.

The more technical section of the book, on new materials, is mainly devoted to the extensive work which has been carried out on the cubic ferrites. A number of these ferrites, of the general formula $MO \cdot Fe_2O_3$, form mixed crystals, and four types of mixed ferrites have been developed, under the trade name "ferroxcube", as core materials suitable for high frequencies. The essential characteristic is a low conductivity combined, for certain mixtures, with a high initial permeability. For a mixed nickel-zinc ferrite, for example, a permeability of 4000 has been obtained. A particular point of interest here is that an enhanced permeability results from the addition to nickel ferrite of zinc ferrite, which is not itself ferromagnetic. Its effect is to lower the Curie point, so bringing the temperature region of high permeability into the working temperature range.

These few examples of the many topics touched on in this book must serve to illustrate its character and content. It is an informal account partly of completed work, and partly of work still in progress and of hypotheses which are being followed up. It should be in the hands of all those interested in ferromagnetism, by whom it will be valued both as a record of achievement under great difficulties, and for its stimulating presentation of new ideas.

E. C. S.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 60, PART 2

1 February 1948

No. 338

The Structure of Photo-sensitive Lead Sulphide and Lead Selenide Deposits and the Effect of Sensitization by Oxygen*

By H. WILMAN

Applied Physical Chemistry Laboratory, Imperial College, London

MS. received 2 June 1947

ABSTRACT. An exploration has been made by electron diffraction and photoconductivity experiments, of the structure of photosensitive lead sulphide and lead selenide deposits prepared by chemical deposition and by sublimation *in vacuo* or in oxygen. In spite of the large variations in photoconductive properties which the deposits show as a result of treatments such as baking in vacuum or oxygen, the PbS and PbSe crystals were found to have a lattice axial dimension which was constant to 0.1 % in all samples measured, even in strongly oxidized deposits in which there was estimated to be 10 to 20 % of the oxidation product. The oxidation product was identified from its diffraction pattern as PbO.PbSO₄. The photoconductivity measurements, however, indicate that the sensitivity is associated with deviations from the stoichiometric compositions PbS and PbSe, and that oxygen treatments increase the sensitivity of PbS to wavelengths between 1 and 3 μ , but do not do so for PbSe.

§1. INTRODUCTION

DURING the war lead sulphide photoconductive cells were manufactured in Germany by Gudden, Kaspar, Kutzscher and others, but the methods of production have not been published. In 1944 methods of manufacture of PbS and PbSe cells were developed at the Admiralty Research Laboratory, Teddington (for details of the vacuum-sublimed deposits see Starkiewicz, Sosnowski and Simpson 1946, Lee and Parker 1946), and a brief account is given below of the structure of the deposits, as revealed by the use of electron-diffraction methods. Electrons of about 60 kv. voltage were used with a camera length of approximately 50 cm. in a Finch type electron-diffraction camera (Finch and Wilman 1937). Early in 1946 further electron-diffraction investigations provided definite information on the amount, chemical nature and form of the oxidation product obtained by heating or subliming (at temperatures above 500° c.) the PbS layers in oxygen at about 0.1 mm. Hg pressure, during their sensitization as photoconductive or photovoltaic cells (Starkiewicz, Sosnowski and Simpson 1946).

The crystal lattice structure and parameters of the PbS and PbSe were practically identical for all the deposits examined; these are stated in §6. Of the photosensitivity data only a brief outline of some general characteristics is given in §6.

* This work was carried out in the Applied Physical Chemistry Laboratory, at Imperial College, London, by arrangement with Professor G. I. Finch, M.B.E., D.Sc., F.R.S.

The vacuum-sublimed deposits were all prepared in well degassed pyrex pumping systems (0.5 to 1 hour degassing at 500° c.) evacuated by mercury diffusion pumps to about 10^{-6} mm Hg, with liquid-air trap to avoid mercury contamination.

§ 2. THE STRUCTURE OF LEAD SULPHIDE AND LEAD SELENIDE DEPOSITS FORMED BY CHEMICAL PRECIPITATION

Deposits of lead sulphide with good photoconductivity prepared by Dr. F. Kicinski at A.R.L. were precipitated as mirror-like layers on pyrex substrates from a solution of lead acetate and thiourea by adding sodium hydroxide solution, and after washing with weak ammonium sulphide solution were dried in air at 100° c. Figure 1 shows an example of the electron-diffraction patterns obtained from such mirror-like slightly translucent deposits, which are dark olive-green by transmitted light. The arc positions show that the PbS crystals tended to be orientated with cube faces, i.e. {001}, parallel to the polished pyrex substrate. Similar patterns, but with less distinct arc development, were given by thicker opaque deposits ($\sim 1\mu$ thick) in which the PbS crystals in the surface layers had only to a very slight extent the preferred {001} orientation. The mean crystal size was estimated to be at least 150 Å. and was in some cases of the order of 500 Å.

Translucent olive-brown PbS deposits were also prepared for comparison with the above, by passing H_2S gas over a lead-acetate solution at room temperature, slightly acidified with acetic acid. These floating mirror-like films were picked up on nickel gauze, transferred to the surface of an H_2S water solution to wash them, removed, and a drop of Aquadag graphite added for lattice-constant reference purposes; they were then dried *in vacuo* in the diffraction camera. The ring-patterns obtained at normal transmission, figure 2, became weakly arced on inclining the film, showing that the PbS crystals had their cubic faces, {001}, more or less parallel to the surface of the solution on which they were formed. The mean crystal diameter was about 250 Å.

Similar mirror-like, translucent reddish-brown layers of PbSe were obtained by passing H_2Se over lead-acetate solution acidified with acetic acid, and these gave patterns like figure 13, where the {110} graphite ring is also present for calculation of lattice dimensions. The PbSe crystals had a mean diameter of about 250 Å. and had tended to grow with {001} planes parallel to the surface of the solution. Such PbSe layers showed no change in diffraction pattern after heating for 30 minutes in air at 100° to 130° c. or in steam at 130° to 150° c.

§ 3 THE STRUCTURE OF LEAD SULPHIDE AND LEAD SELENIDE DEPOSITS PREPARED BY VACUUM SUBLIMATION

Some of the early experimental PbS cells made by the vacuum sublimation method were cut open and immediately re-evacuated in the electron-diffraction camera, and an electron beam allowed to graze the inner cylindrical surface of the deposit on the electrode region of the pyrex cell wall. Three types of deposit were examined: (a) PbS volatilized and condensed *in vacuo* then superficially acted upon by sulphur vapour at low pressure and temperature about 500° c. (figure 3); (b) PbS volatilized and condensed in presence of sulphur vapour (figure 4); (c), as for (b), but resublimed in oxygen at 1 mm. Hg pressure.

In all these the mean crystal diameter of the PbS was estimated to be between 150 and 500 Å. or more, and there was usually a marked tendency for the PbS

crystals to grow in a one-degree orientation, either {111}, {110}, {210}, or {311}; in some cases two, or even three, of these were present in the same layer. For example, figure 4 shows that {311} planes in some crystals and {210} in others were parallel to the pyrex substrate. The type of orientation developed presumably depends on the temperature of the surface and the rate of condensation. The one deposit of the third type examined had a larger crystal size than the others, 500 Å. or more, as shown by the incompletely formed spotty rings, cf. figures 5-8. This was to be expected, for the temperature of sublimation of the PbS was higher ($\sim 580^\circ\text{C.}$) in oxygen than *in vacuo* ($450^\circ\text{--}500^\circ\text{C.}$), and since that of the substrate was also higher, there was more rapid aggregation of the PbS to form larger crystals. Two exceedingly faint unidentified rings were present in this pattern in addition to the PbS rings.

Seven vacuum-sublimed translucent orange-brown PbSe deposits in experimental cells were examined. The PbSe used was prepared by passing H_2Se into lead-acetate solution followed by much repeated washing of the precipitate and drying at about 100° to 130°C. All these deposits yielded clear ring or arc patterns, as in figure 14, due to PbSe crystals either randomly disposed or in strong one-degree {001} orientation on the substrate. The estimated mean crystal size varied between about 150 Å. and 500 Å. or more, increasing with increasing duration and temperature of the vacuum bakes which the layer had undergone after the initial condensation. In one case where a cell had developed a noticeable leak at one of the glass-tungsten seals during formation of the layer, which had become grey in colour, an unidentified diffraction pattern was obtained (figure 16, table 4). This presumably corresponds to one or more of the compounds of Pb, Se and O.

§ 4. THE STRUCTURE OF LEAD SULPHIDE DEPOSITS SENSITIZED BY SUBLIMING AND BAKING IN OXYGEN

(i) Photoconductive deposits

Two cells A and B (made by J. Starkiewicz at A.R.I.) which had undergone the normal type of activation treatments by slow sublimation and further baking in oxygen, yielded the electron diffraction patterns, figures 5 and 6; net-plane spacings are given in table 1. In both cells the pumping stem had a narrow constriction impeding escape of SO_2 gas to the liquid air trap, and the final (modulated-light) sensitivity reached the average good high level attained with most cells of this type. Both PbS deposits had a matt grey-black surface and were opaque.

The patterns of spotty arcs (figures 5 and 6) showed that the PbS crystals had relatively large size, probably ~ 1000 Å. diameter, and had grown so that some had {110} planes parallel to the pyrex substrate, others {001}. The lattice dimension was $a = 5.917$ kx. relative to graphite $a = 2.456_3$ kx. The additional fainter arcs in the patterns showed the presence of an oxidation product (forming about 10-20% of the diffracting layer) which was also crystalline, with mean crystal diameter of the order of 500 Å., and was also (especially in cell A) strongly orientated, the net-plane parallel to the substrate having a spacing of 1.228 kx. This oxidation product was identified from its net-plane spacings as PbO.PbSO_4 (see (iii) below), the crystal habit apparently varying according to the conditions of oxidation and causing some variation in ring intensities, without much change in the lattice dimensions and net-plane spacings.

(a) Lanarkite. PbO.PbSO ₄ calculated		(b) PbS heated in air 10 min.; 350° C. (figure 9)		(c) PbS heated in air 7 min. to 550° C. + 5 min. at 550° C. (figure 10)		(a) Lanarkite PbO.PbSO ₄ calculated		(b) PbS heated in air 10 min.; 350° C. (figure 9)		(c) PbS heated in air 7 min. to 550° C. + 5 min. at 550° C. (figure 10)	
<i>d</i> (kx.)	Indices	<i>d</i> (kx.)*	Intensity	<i>d</i> (kx.)*	Intensity	<i>d</i> (kx.)	Indices	<i>d</i> (kx.)	Intensity	<i>d</i> (kx.)*	Intensity
—	—	8.21	MF	—	—	1.661	132	1.650	MF	1.646	F
6.40	001	7.31	F	—	—	1.600	004	1.591	MF	1.601	MF
6.16	200	6.409	F	6.24	MF	1.587	132	—	M	—	—
5.92	201	5.878	F	5.73	F	1.566	313	—	MF	1.552	MF
5.33	204	5.234	F	5.159	MF	1.551	313	1.555	MF	—	—
5.16	110	—	F	4.704	VF	1.544	422	1.510	F	1.504	MF
4.42	111	4.447	F	4.415	MF	1.497	424	1.480	F	1.476	MF
—	—	4.178	VF	—	—	1.455	002	—	F	—	—
3.68	111, 201	3.696	MF	3.666	MF	1.450	133	1.445	F	1.442	F
3.52	302	—	F	3.495	M	1.432	624	1.426	F	1.418	F
3.31	311	3.491	F	3.430	S	1.420	040	—	—	—	—
3.33	310	3.333	MI	3.310	S	1.385	024	1.390	F	1.382	MF
3.20	002	3.258	S	3.236	S	1.384	240	1.359	F	1.355	MF
3.08	400	3.078	MF	—	—	1.357	133	1.330	VF	1.332	MF
2.96	302	2.946	S	2.918	S	—	—	1.312	VF	1.305	VF
2.84	020	2.845	S	2.836	S	1.317	242	—	—	—	—
2.665	404	2.659	MI	—	—	1.296	042	—	—	—	—
2.59	311	2.571	MI	2.566	MF	1.294	622	—	—	—	—
2.58	220	2.462	F	2.463	F	1.290	440	—	—	—	—
2.42	202	2.416	F	2.419	F	1.287	314	—	—	—	—
2.38	113	2.334	F	2.324	MF	1.283	442	1.281	MF	1.279	F
2.23	602	2.256	M	2.242	MS	1.252	134	1.260	MF	1.249	VF
2.21	322	2.207	F	—	—	—	—	—	—	—	—
2.17	313	—	—	2.169	VF	1.227	333	1.228	?	—	?
2.15	204	—	—	—	—	1.226	603	(graphite 110 diffraction)	—	1.228	?
2.12	022	2.117	MI	2.108	MF	1.224	242	1.209	VF	1.205	VF
2.09	420	—	—	—	—	1.209	404	1.192	F	1.192	VF
2.05	600, 122	2.030	MS	2.040	MS	1.198	642	—	—	—	—
1.972	603	—	—	—	—	1.183	10, 0, 5	—	—	—	—
1.958	113	1.956	S	1.960	F	1.174	606	1.172	VF	—	—
1.870	130	1.850	F	1.898	F	1.171	134	—	—	—	—
1.841	222	—	F	—	S	1.139	226	—	—	1.172	VF
1.840	402	—	—	1.837	—	1.128	206	—	—	—	—
1.830	131	—	S	—	—	1.126	442	1.131	F	1.128	MF
1.777	6, 0, 12	1.829	—	—	—	1.115	15, 0, 4	1.120	VF	—	—
1.764	131	—	—	—	—	1.112	424	—	—	—	—
1.762	404	—	—	—	—	1.109	903	—	—	—	—
1.755	622	—	—	—	—	1.106	444	1.103	VF	—	—
1.744	331	—	—	1.734	MF	1.098	244	—	—	1.095	MF
1.735	314	—	—	—	—	—	—	1.085	VF	—	—
1.720	330	—	—	1.720	MI	—	—	1.073	VF	—	—
—	—	1.685	MI	—	—	—	—	—	—	1.077	MF
1.663	620	1.667	F	1.663	MI	—	—	—	—	—	—

Wolfe (1938). The latter authors give the size of the monoclinic pseudohexagonal unit cell as $a = 13.73 \text{ \AA}$, $b = 5.68 \text{ \AA}$, $c = 7.07 \text{ \AA}$, and $\beta = 116^\circ 13'$. The cell is base-centred, thus only hkl diffractions appear which have $(h + k)$ even. A calculation of the net-plane spacings corresponding to possible diffractions gave the values in table 2, which includes all values above 2 \AA . and many below. Relative intensities were obtained by the experiments described in § 5, below.

In the comparison of the lists of spacings and diffraction intensities in tables 1 and 2 the following points must be realized: (i) owing to the orientation of the crystals in the photocell deposits, not all the diffractions in table 2 will be visible above the shadow edge in the reflection patterns; (ii) the relative intensities of the arcs which do appear will be different from those in table 2, owing to the orientation and to absorption which weakens beams emerging near the shadow edge; (iii) planes which are parallel to the surface in most of the crystals give especially strong arcs whose centres lie on the plane of incidence; (iv) there may be a slight difference in the lattice dimensions and the relative diffraction intensities such as that observed in the patterns from PbS heated in air at 350°C . and 550°C . (table 2), although the net-plane spacings are all standardized by comparison with the graphite 110 spacing; (v) strong PbS diffraction arcs may obscure other weaker arcs in their vicinity; (vi) smallness of the radii and of refraction effects decrease the accuracy of measurement of the larger net-plane spacings.

Bearing in mind the above conditions the oxidation product in the photocell deposits was identified from the spacings given in table 1 and from the circumferential positions of the arcs in the patterns, figures 5 to 8, as $\text{PbO} \cdot \text{PbSO}_4$ (or approximating closely to that composition and lattice structure) for the following reasons:

(i) The spacings in the first four lists in table 1 agree well with themselves and with many of those in table 2, notwithstanding the faintness of some of the diffractions. (ii) The lanarkite structure accounts for the large spacings of the two innermost diffractions of figure 8 (cell A). (iii) On the basis of the lanarkite structure, the net-planes which are orientated parallel to the substrate are densely populated planes having small indices, $\{111\}$ in cells A, B and D, and $\{111\}$ and $\{102\}$ together in cell D; the second and third orders of reflection from the $\{111\}$ planes give as the $\{111\}$ spacing the value 3.68_4 \AA . in agreement with the value for lanarkite in table 2. (iv) The circumferential positions of the arcs on the other rings appeared to correspond well with those to be expected from lanarkite crystals in these orientations. (v) The plane spacings of the transmission patterns obtained after complete oxidation of PbS films in air at 350°C . and 550°C . agreed closely with those calculated for lanarkite (table 2).

In order to obtain diffraction patterns from other Pb-O-S compounds several other materials were examined by electron diffraction, and though they did not correspond to that of the oxidation product discussed here, they are of some interest and provide another example of appreciable changes of diffraction intensities associated with only slight changes in lattice dimensions. Figure 11 and table 3 (spacings calibrated by graphite) show the diffraction data obtained by transmission from a floating white surface layer precipitated by passing SO_2 gas over lead-acetate solution. This is certainly not PbSO_4 , but the pattern seems to have some similarities with those of the 2-, 3- and 4-PbO.PbSO₄ oxysulphates. When this material was heated in air for 5 minutes at 300°C . it gave a pattern (see table 3), in which the ring intensities were appreciably changed though their

relative radii were nearly unaltered; after a further 45 minutes in air at 550° c. it gave figure 12 (table 3), which may possibly be closely related to the x-ray pattern from $2\text{PbO} \cdot \text{PbSO}_4$ though it is not identical with this.

Table 3. (a) and (b) are nearly identical and are closely related to the pattern (b) of table 2; the spacings of (c) correspond with most of those of table 2 (c) but the intensities are different.

(a) White pp. from Pb acetate + SO_2		(b) Pp. of (a) heated 5 min. at 300° c. in air		(c) Spec. of (b) heated 45 min. at 550° c. in air	
d (kx.)*	Intensity	d (kx.)*	Intensity	d (kx.)*	Intensity
8.07	VVF	—	—	—	—
6.75	VF	6.4	VF	—	—
5.182	F	5.14	VF	4.45	F(D)
4.261	M	4.24	MF	4.11	F(D)
3.753	MS	3.73	MS	3.68	MF(D)
3.414	MF	3.40	MF	3.48	F(D)
3.218	MS	3.215	M	3.286	MF
3.122	MS	3.110	M	2.963	VS
2.905	MS	2.905	S	2.849	S
2.723	S	2.719	S	(2.63)	VF
2.576	F	2.575	VF	2.528	M
2.469	MF	2.444	F	2.395	MF
2.341	VF	—	—	—	—
2.248	M	2.246	F	2.263	F
2.136†	MF	2.127†	MF	2.142	F
2.058	MS	2.064	M	2.047	M
2.025‡	MS	2.025‡	M	—	—
1.958	MF	1.960	F	1.954	F(D)
1.903	MF	1.901	F	1.896	F(D)
1.865	F	1.852	F	1.843	F(D)
1.831	F	1.825	VF	—	—
1.760	VF	—	—	1.754	F(D)
1.719	MF	1.715	MF	1.716	F(D)
1.704	F	—	—	—	—
1.659	F	1.657	F	1.669	F(D)
1.644	F	—	—	—	—
1.606	MF	1.613	F	1.615	F(D)
1.571	MF	1.566	F	—	—
1.502	F	1.502	VF	1.520	F(D)
1.457	M	1.454	MF	1.462	M(D)
1.446	M	—	—	1.419	VF
1.393	MF	1.391	VF	1.386	VF
1.368	F	—	—	1.352	VF
1.339	F	—	—	—	—
1.318	VF	—	—	1.311	MF(D)
1.305	VF	—	—	—	—
1.283	VF	—	—	1.289	MF(D)
1.254	M	1.253	MF	1.249	VF(D)
1.205	VF	—	—	1.206	VF(D)
1.189	VF	—	—	1.169	VF(D)

* Relative to graphite, d_{110} 1.228 kx.

† May be entirely the graphite 100 ring.

‡ May be partly but only in small part due to graphite 101 diffraction.

§ 5 THE OXIDATION OF PbS AND PbSe IN AIR AT ATMOSPHERIC PRESSURE

Translucent olive-green films of PbS with strong metallic reflexion were formed by passing H_2S over lead nitrate solution. After washing by floating on distilled water these gave normal PbS diffraction patterns like figure 2. When such layers were heated in air at atmospheric pressure at 350–370° c. for 10 minutes



Figure 1. Chemically deposited PbS layer (good sensitivity)



Figure 2. PbS layer from H_2S over Pb acetate solution.



Figure 3. PbS cell; layer formed by vacuum sublimation and treatment with S

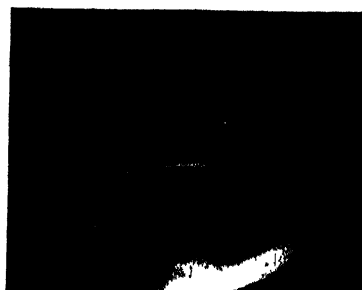


Figure 4. PbS cell; layer formed by vacuum sublimation in presence of S.



Figure 5. PbS cell "A"; layer sublimed in O_2 .



Figure 6. PbS cell "B"; layer sublimed in O_2 .



Figure 7. PbS cell "C"; formed in oxygen, but photovoltaic type.



Figure 8. PbS cell "D"; formed in O_2 , but photovoltaic type.

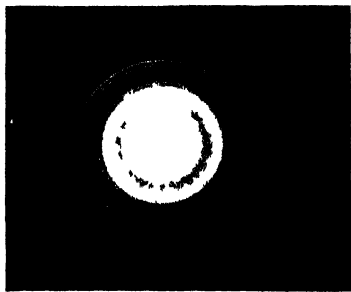


Figure 9 PbO PbSO₄ by heating PbS in air at 350 °C for 10 minutes (Two outermost rings graphite)

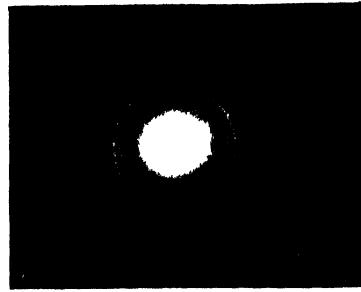


Figure 10 PbO PbSO₄ by heating PbS in air at 550 °C for 5 minutes (Graphite present as reference)



Figure 11 White layer by SO over Pb acetate solution

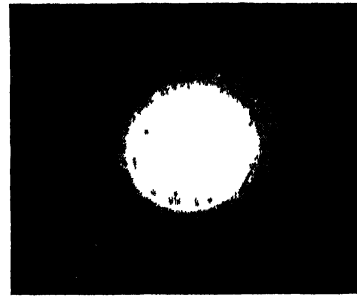


Figure 12 Deposit of figure 11 heated in air at 550 °C for 45 minutes

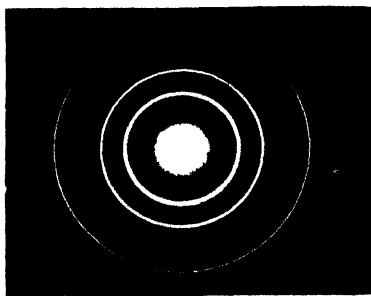


Figure 13 PbSe layer from H₂Se over Pb acetate solution

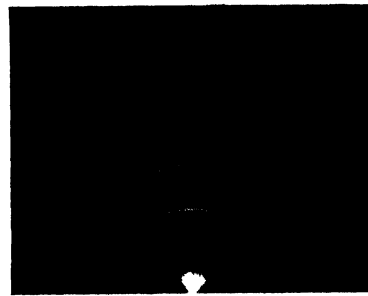


Figure 14 PbSe cell layer formed by vacuum sublimation

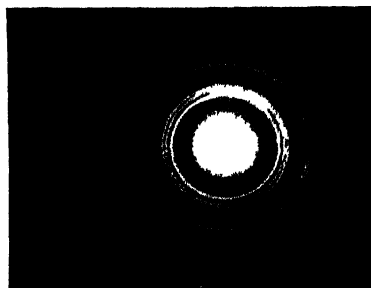


Figure 15 PbSe heated in air at 350 °C for 10 minutes



Figure 16 PbSe cell after a leak had caused oxidation

they became transparent yellowish-grey, giving a different pattern, figure 9, which corresponds exactly to PbO.PbSO_4 . Table 2 shows the plane spacings (calibrated by comparison with graphite) and the relative intensities of the diffraction rings, and these are compared there with the data for lanarkite, PbO.PbSO_4 .

A fresh PbS layer, heated in air up to 550°C . for 7 minutes and kept there for 5 minutes, became milky-transparent and gave a pattern, figure 10 and table 2, with spacings nearly identical with those of figure 9, but with a difference in the relative intensities of the corresponding rings, the random disposition of the crystals in both cases was checked by photographs with the specimen inclined to detect any arcing of the rings. This slight change in lattice dimensions may indicate a slightly different degree of oxidation in the two cases with only a small difference in crystal lattice dimensions. The change in relative intensities of the rings may be attributed to a difference in the external shape of the crystals, which would cause different absorption of the various diffracted beams passing through different thicknesses of crystal in different directions (Böhm and Gantner 1928, Fordham 1940, Brandenberger 1945).

Similar PbSe layers formed by passing H_2Se gas over lead-acetate solution acidified with acetic acid gave unchanged ring patterns by transmission electron diffraction after heating the layers at 100 to 130°C . in air for 30 minutes or in steam at 130 to 150°C . for 30 minutes; but heating in air at 350°C . for 10 minutes converted them completely to a transparent yellowish oxidation product (figure 15), the diffraction data for which are given in table 4. Comparison with the pattern data from the PbSe layer oxidized by a leak in a cell during a vacuum bake (see §3) seems to show some similarities in spacings though differences in ring intensities.

§6. SUMMARY AND DISCUSSION OF RESULTS

A. Electron-diffraction results

The following results were obtained by electron diffraction:

1. All the lead sulphide and lead selenide deposits (~ 500 to $10,000 \text{ \AA}$. thick) which were examined by electron diffraction were found to be crystalline with no noticeable signs of any amorphous material; diffraction methods would only detect proportions of the order of 1% or more of the material.
2. All the deposits had the face-centred-cubic structure of sodium chloride type, previously found for PbS and PbSe by x-ray diffraction (*Strukturbericht, Z. Kristallogr.*, 1, 1928).
3. The mean crystal size and the extent and nature of the orientation present in these polycrystalline deposits varied widely from one specimen to another.
4. All the PbS deposits showed an absence of appreciable variation of the lattice dimensions within the crystals of each deposit, and all the deposits examined had the same lattice dimensions within the limits of accuracy of the measurement (to approximately 0.1%), whether the deposits had been prepared by (a) chemical deposition from a lead acetate + thiourea + sodium hydroxide reaction on to pyrex substrates, (b) as a floating mirror-like layer by passing H_2S gas over concentrated slightly acidified lead acetate or lead nitrate solution, the film being washed by floating on an H_2S -water solution and dried in vacuum at room temperature, (c) sublimation of well-washed and dried precipitated PbS , in an evacuated pyrex cell at about 10^{-6} mm. Hg pressure, with or without further short bakes in vacuum

Table 4.

(a) PbSe ($a = 6.117 \text{ k}\text{\AA}$) ¹			(b) PbSe heated in air 1 hour at 350° c. (figure 15)		(c) PbSe cell which developed a leak during a vacuum bake (figure 16)	
$d (\text{k}\text{\AA})$	Intensity	hkl	$d (\text{k}\text{\AA})^*$	Intensity	$d (\text{k}\text{\AA})^\dagger$	Intensity
			6.16	MF		—
			5.34	F	4.67	F
			4.45	F(10)	4.33	F
			3.92	MF	4.06	F
			3.68	MF	3.69	MF
3.532	M	111	3.515	M	3.413	F
3.058	M	200	3.262	S(B)	3.253	F
					3.068	MF
			2.970	MS(B)	2.988	VF
			2.845	MF	2.838	MF
			2.763	MS	2.711	S
			2.578	MF		—
			2.506	F	2.468	F
			2.404	F	—	—
			2.291	MF	2.309	MS
2.163	S	220	2.131†	M	2.191	M
			2.063	M	2.110	MF
			1.992	MF	2.009	F
			1.946	MF	1.924	F
			1.906	F	—	—
1.844	MF	311	1.837	MF	1.841	F
1.766	MS	222	1.764	MF	1.764	MS
			1.702	MF	1.706	F
			1.678	MF	1.667	F
			1.634	M	1.627	F
1.529	VF	400	1.588	F	1.570	M
			1.529	MF	1.540	MF
				—	(1.520)	VF
			1.473	M	1.465	F
1.403	MF	331	1.424	F	1.425	F
1.368	S	420	1.327	VF	1.352	MS
			1.304	MF	1.296	MS
			1.287	MF	1.269	VF
1.248	S	422	1.254	F	1.238	M
1.177	F	$\left\{ \begin{array}{l} 333 \\ 511 \end{array} \right.$	—	—	1.172	M
					—	—
					1.149	VF
					1.127	VF
					1.097	VF
					1.076	M
					1.023	M

* Relative to graphite $d_{110} = 1.228 \text{ k}\text{\AA}$.

† Not relative to graphite; absolute values as a whole may be a few % in error.

‡ May be entirely due to the graphite 100 diffraction ring.

at between 200° and 400° c., (d) sublimation in oxygen at 0.1 to 0.3 mm. Hg pressure at 500° to 580° c.

A similar constancy (to $\sim 0.1\%$) was also found for PbSe deposits prepared (a) as a floating mirror-like layer by passing H_2Se gas (from either $\text{FeSe} + \text{dilute HCl}$, or $\text{Al}_2\text{Se}_3 + \text{boiled distilled water}$) over concentrated lead acetate solution acidified with acetic acid, (b) by method (a) followed by heating in air at 130° to 150° c. for 30 minutes, (c) by method (a) followed by heating in steam at 150° c. for 30 minutes, (d) by sublimation on to pyrex in a cell evacuated to 10^{-6} mm. Hg, with or without further short bakes (of the order of 10–30 minutes) in vacuum at 200° to 400° c.

The mean lattice constants found for PbS and PbSe were $a = 5.917 \pm 0.005$ kx., and 6.117 ± 0.005 kx. respectively*, relative to $a = 2.456_8$ kx. for graphite (Trzebiatowski 1937, Nelson and Riley 1945, Finch and Fordham 1936, Finch and Wilman 1936, cf. also Wilman 1940), the 110 graphite diffraction ring with $d = 1.228$ kx. being used as the comparison standard.

Previous x-ray measurements, mainly for the more or less impure minerals galena (PbS) and clausthalite (PbSe) have given the following lattice dimensions: for PbS, $a = 5.96_{\pm 4}$ Å. (Lehmann 1924), 5.91 Å. (Kolderup 1924-5), for very pure galena from Příbram Bohemia, 5.93 Å. (Ramsdell 1925), 5.95 Å. (Goldschmidt 1927-28), 5.96 Å. (Bravo 1926), 5.91 ± 0.03 Å. (Vlasak and Trousil 1934) 5.935 Å. (von Zeipfel 1935 a, corrected for more accurate value of λ of Al K α x-rays by S. von Friesen 1935), 5.92334 kx. (von Zeipfel 1935 b), 5.94 Å. (Hanawalt, Rinn and Frevel 1938), 5.92 Å. and 5.93 Å. for galena containing 6% Bi (Goldschmidt 1927-28), for PbSe, $a = 6.14$ Å. (Ramsdell 1925), 6.162 Å. (von Olshausen 1925); and $6.13_{\pm 1.5}$ Å. from Pb-Se melted together in N₂ (Goldschmidt 1927-28).

The constancy of the lattice axial dimension to within 0.1% for PbS and PbSe in the various degrees of oxidation can be taken as showing that the proportion of excess Pb atoms or S defect ("holes"), or of S or Se respectively replaced by O ions in the PbS or PbSe lattice is very small, and probably less than about 1%, even in the strongly oxidized deposits in which up to 10 to 20% of PbO. PbSO₄ has segregated out. For example, if the ionic radii of Pb⁺⁺, S⁻ and O⁻ are 1.32, 1.74 and 1.32 Å. respectively, then replacement of 1% of the S⁻ by O⁻ would be expected to result in a reduction in the lattice dimensions of the PbS by approximately 0.08%. However, presence of excess Pb or of "holes" due to missing S⁻ ions might tend to increase the lattice dimension and counterbalance a reduction due to presence of O⁻ ions. Since the outer diffraction rings are not noticeably less sharp than the inner rings, however, it seems that lattice dimension variations larger than 0.1% do not occur throughout the PbS crystals of 500 to 1000 Å. diameter, or occur only locally to a very small extent.

5. The fact that well-defined diffraction rings were obtained from the photo-sensitive deposits by the "reflection" method shows that these deposits were relatively rough even when they appeared to be highly reflecting and optically flat, the height of the "peaks and valleys" being at least 150 Å., and mostly 500 Å. or more in the vacuum-sublimed deposits. Correspondingly the electron-diffraction patterns show the structure of the deposit to this depth below the outer parts of the surface.

6. While the electron-diffraction patterns from the chemically-prepared, and also the vacuum-sublimed, PbS and PbSe deposits showed no trace of any other material present, those from PbS sublimed or baked in oxygen at ~ 0.1 mm. Hg pressure (the SO₂ formed being trapped by liquid air cooling) showed besides the strong PbS rings or arcs a much fainter pattern due to an oxidation product, which was estimated to be present in the surface layers to the extent of about 10-20%. When the oxidation was slight this pattern was represented by only one or two very faint diffuse rings, but in the four typical cells of this type examined, two of photo-conductive and two of photovoltaic type, it consisted mainly of clearly defined rings

* This can be converted to Ångströms by multiplying by 1.00202, the factor recently agreed on by the X-ray Analysis Group of the Institute of Physics and the American Society for X-ray and Electron Diffraction.

or sharp arcs, from the positions of which the oxidation product was identified as lead oxysulphate (lanarkite), $\text{PbO} \cdot \text{PbSO}_4$. Translucent PbS layers heated in atmospheric air at 350°C. for 10 minutes were completely converted to a pale yellow transparent material also identified as $\text{PbO} \cdot \text{PbSO}_4$; and heating at 550°C. in air, followed by cooling to room temperature, caused only very slight changes in the lattice dimensions but distinct changes in the relative intensities of the diffraction rings, probably due to a different crystal shape having been developed, causing greater absorption for some diffracted beams than for others.

Lead selenide deposits, when similarly heated or sublimed in oxygen, were also converted into a greyish-yellow oxidation product; this was not identified owing to lack of diffraction data for the lead oxyselenium compounds. No improvement in photosensitivity was obtained from these oxidizing treatments of PbSe.

B. Photoconductivity results

With regard to the photosensitivity of the PbS and PbSe cells, only the following results will be stated here for consideration with the diffraction results, and a more detailed account of PbS cells will be published shortly by Starkiewicz *et al.*

1. The photoconductivity properties of the deposits varied widely according to the conditions of formation of the deposit (e.g. in vacuum-prepared cells the temperature, rate of condensation, thickness of deposit, pressure of residual gas in the apparatus, nature of substrate etc.), and the further heat treatments in vacuum or in oxygen. No definite relationship was observed between the magnitude or the spectral distribution of the photosensitivity and the crystal size or kind and degree of orientation of the PbS or PbSe as shown by the diffraction electron patterns. Since no full and systematic investigation has yet been made, only brief general indications will be stated here about the photoconductive and other properties of the deposits. It was noticeable, however, that the sensitization of PbS cells by subliming the material (or baking it) in oxygen led to a large increase in the crystal size as shown by the incomplete formation (spottiness) of the sharp diffraction rings due to the small number of crystals in the path of the electron beam. It is well known that crystal growth is much faster in the presence of oxygen than in vacuum, though it is not clear whether such accelerated growth is accompanied by appreciable diffusion of oxygen into the crystal lattice in solid solution. In the present case the sensitization in oxygen was found (Lee and Parker 1946) to be associated with development of a peak in the spectral sensitivity curve at a wavelength near 2.7μ in addition to that at 1.1μ , evidently indicating strong adsorption or absorption of oxygen. The relation of the rise in infra-red sensitivity to the observed increase in crystal size and the formation of relatively large crystals of $\text{PbO} \cdot \text{PbSO}_4$ is not yet clear, but the state of over-oxidation evidently corresponds to segregation of a large proportion of $\text{PbO} \cdot \text{PbSO}_4$ among the PbS crystals.

The observed absence of beneficial effects of oxygen treatments on the photosensitivity of PbSe deposits is probably due to the absence of appreciable "reducing" action, i.e. liberation of lead, and to too rapid segregation of the oxidation product. Segregation is accelerated in this case because selenium is not readily removable as SeO_2 , which is a solid at temperatures up to several hundred degrees Centigrade, whereas SO_2 is gaseous and is removed from the layer.

2. As a brief general indication of the electrical properties of the deposits it may be stated here that the resistances of the PbS and PbSe deposits between graphite or platinum electrodes 1 cm. long and 1 mm. apart were in the range 10 to 1000 k Ω and were approximately doubled when the cells were cooled in liquid air. The PbS cells, when illuminated by a broad, roughly collimated beam of unmodulated white light from a 36-watt lamp, showed a decrease of resistance of up to about half the dark resistance and a ratio of signal to noise of unity was observed with the cells exposed to 800 cycles/sec. modulated radiation from a 3 mm. aperture in front of a black body at 200° c., at a distance of the order of 70 cm. Further details of the sensitivity are given by Lec and Parker (1946). The PbSe cells showed only negligible photoconductivity ($\sim 2\%$ change in resistance) with continuous illumination, and with the modulated radiation had mainly a sensitivity of the order of 1–2% of that of good PbS cells.

3. In spite of the observed degree of constancy of the lattice dimensions calculated from the electron-diffraction patterns (§4) obtained from cells whose properties differed widely, the electrical properties of both the PbS and PbSe deposits, measured at room temperature, varied within a wide range when previously well-outgassed cells were baked (at between 200° and 500° c.) while evacuated to about 10^{-6} mm. Hg pressure. In the case of PbSe, for example, the translucent reddish-orange deposits made by vacuum sublimation in presence of excess selenium, had low resistance ($\sim 10\,000\ \Omega$), very small "noise" and also very low signal and signal/noise ratios, when using a small red-hot tungsten filament to give a light beam which was interrupted 800 times/sec. by a rotating slotted disc, the resulting alternating current being amplified. Successive periods of 10 minutes heating in vacuum at about 350° c. (Se begins to vaporize appreciably in vacuum at about 200° c.) then caused progressive increase of the resistance, noise, signal and signal/noise ratio, as in figure 17, until a maximum was reached for all these properties nearly simultaneously. By driving in excess selenium and then again heating *in vacuo* in successive stages, these properties again passed through a maximum and returned to a nearly constant low level again (figure 17). (Note that in figure 17 the periods of baking stated were at the temperatures indicated by the apexes of the triangles in the lower diagram, additional to the 4–5 minute heating-up.) These observations, together with the visible loss of selenium from the cell to form a condensed deposit in the cooler parts of the pumping stem, appear to indicate that the vacuum baking resulted in at least the upper layers of the PbSe deposit losing selenium and passing from a composition having excess Se, through the stoichiometric composition PbSe with no Pb or Se impurity centres which corresponds to the maximum of resistance etc., to one with a defect of Se (i.e. excess Pb). In view of these very wide variations the constancy of the lattice dimensions is remarkable. Wide variations in electrical properties of PbS layers have been described by Hinterburger (1942) and of PbSe layers by Eckart and Raithel (1941). The progressive decrease of the successive maxima in figure 17 may possibly be due to gradual oxidation resulting from the minute trace of residual air in the vacuum system or oxygen evolved from the selenium side tube, or it may be due to some other cause such as loss of selenium from the lower parts of the layer so as to make it tend towards stoichiometric composition (and later recede from it again due to Se defect?), or perhaps a gradual aggregation to form larger more isolated PbSe crystals owing to the energetic diffusion of the selenium in the layer. Figure 18, plotted from these same measurements shows a progressive

increase in resistance of the deposit for a given signal/noise ratio. The initial line 5 6 7-8 passes near the origin, corresponding to a relation of form $(S/N) = Ae^{BR}$. Similar conditions existed for vacuum heat-treatment of PbS cells. Heat treatment of PbS or PbSe in oxygen resulted in more complex variations which will not be described here.

C. General summary

To sum up all the above results, the preliminary electron-diffraction observations have provided a very much more definite knowledge of the nature of the photosensitive PbS and PbSe layers than has hitherto been available. It has identified the composition and defined the form of the crystalline oxidation

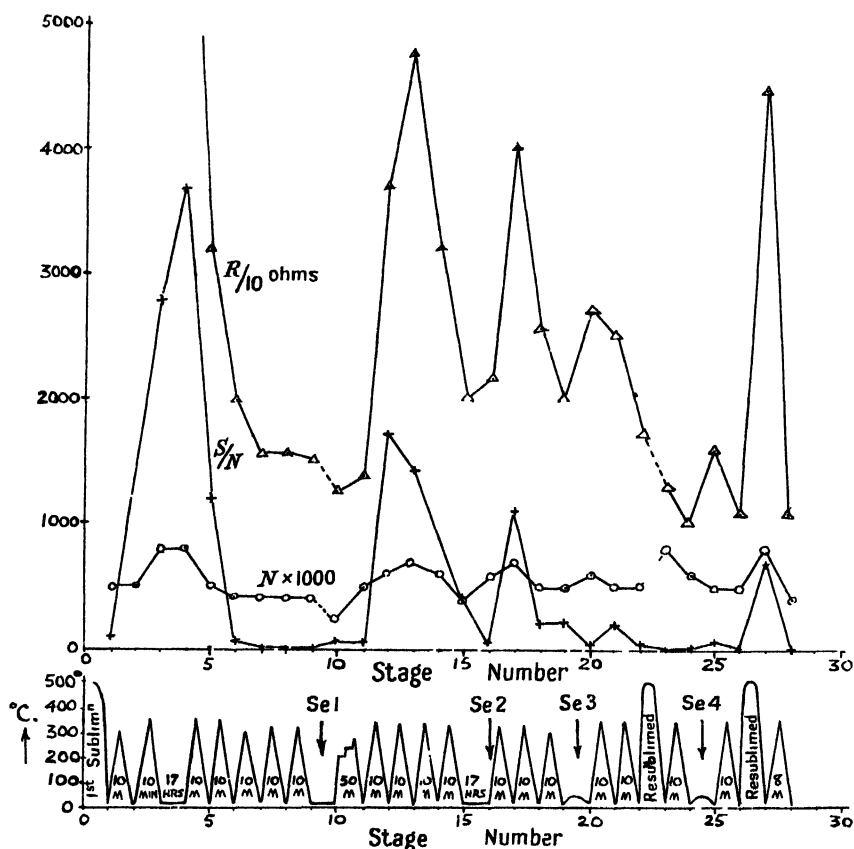


Figure 17.

product formed during sensitization of PbS layers in oxygen, and in conjunction with photoconductivity measurements it has indicated that the crystal orientation has little effect on the photosensitivity compared with other factors such as, especially, very slight deviations from stoichiometric composition, the size of the crystals in the layer, and probably the distribution of the $PbO \cdot PbSO_4$ among the PbS crystals. The axial length of the PbS crystals was $a = 5.917$ kx. and that of PbSe was 6.117 kx., and no deviation greater than 0.1% occurred whether the PbS contained excess S or excess Pb (as indicated by the photoconductive properties), or whether it was unoxidized or relatively heavily oxidized.

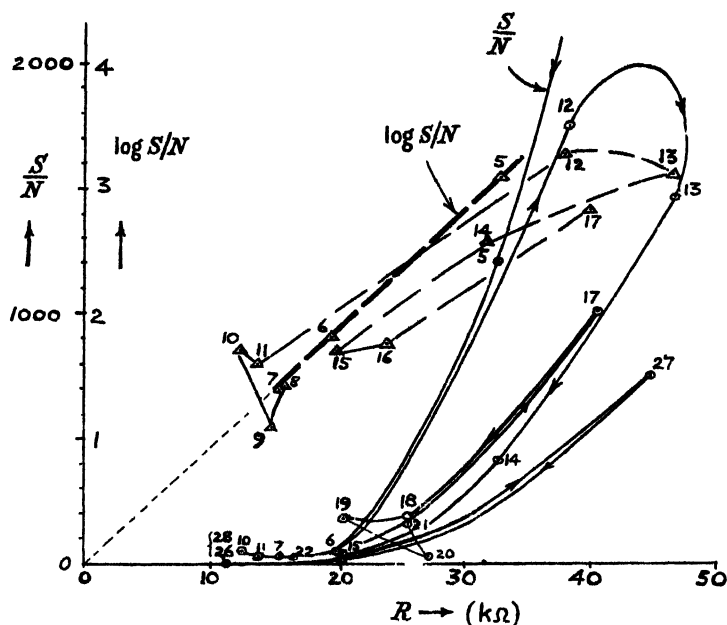


Figure 18.

ACKNOWLEDGMENTS

The author wishes to thank the Admiralty for permission to publish these results, Captain J. Starkiewicz and Dr. F. Kicinski for making available typical PbS cells for the electron-diffraction examination, and Messrs. R. Jennings and B. W. Soole for their assistance in experiments on the photosensitive properties of PbSe cells.

REFERENCES

- VAN ARKEL, 1925, *Physica*, **5**, 162.
 BARONI, A., 1938, *Gazzetta*, **68**, 387.
 BASCHE, W., and MARK, H., 1924, *Z. Kristallogr.*, **64**, 1.
 BÖHM, J., and GANTNER, F., 1928, *Z. Kristallogr.*, **69**, 17.
 BRANDENBERGER, E., 1945, *Röntgenographisch-analytische Chemie*, p. 130 (Verlag Birkhäuser, Basel).
 BRAVO, F. M., 1926, *Anales Soc. Española Fis. Quim.*, **24**, 611.
 BYSTRÖM, A., 1943, *Arkiv Kemi. Min. Geol.*, **17 B**, No. 8.
 BYSTRÖM, A., 1944, *Arkiv Kemi. Min. Geol.*, **18 B**, No. 10.
 BYSTRÖM, A., 1945, *Arkiv Kemi. Min. Geol.*, **20 A**, No. 11.
 BYSTRÖM, A., and WESTGREN, A., 1943, *Arkiv Kemi. Min. Geol.*, **16 B**, No. 14; *Chem. Zentr.*, **2**, 1076.
 CLARK, G. L., 1940, *Applied X-Rays*, p. 31 (McGraw-Hill, New York and London, 3rd ed.).
 CLARK, G. L., MRGUDICH, J. N., and SCHIELTZ, N. C., 1936, *Z. anorg. Chem.*, **229**, 401.
 CLARK, G. L., and ROWAN, R., 1941, *J. Amer. Chem. Soc.*, **63**, 1305 and 1932.
 CLARK, G. L., and TYLFR, W. P., 1939, *J. Amer. Chem. Soc.*, **61**, 58.
 CLARK, G. L., SCHIELTZ, N. C., and QUIRKE, T. T., 1937, *J. Amer. Chem. Soc.*, **59**, 2305.
 DARBYSHIRE, J. A., 1932, *J. Chem. Soc.*, 111.
 DAVIDSON, H. R., 1941, *Amer. Mineral*, **26**, 18.
 DICKINSON, R. G., and FRIATUF, J. B., 1924, *J. Amer. Chem. Soc.*, **46**, 2457.
 ECKART, F., and RAITHEL, K., 1941, *Naturwissenschaften*, **29**, 572.
 FINCH, G. I., and FORDHAM, S., 1936, *Proc. Phys. Soc.*, **48**, 85.
 FINCH, G. I., and WILMAN, H., 1937, *Ergebn. exakt. Naturw.*, **16**, 353.
 FORDHAM, S., 1940, *Nature, Lond.*, **146**, 807.

- VON FRIESEN, S., 1935, *Uppsala Universitets Arsskrift*, p. 14.
 GOLDSCHMIDT, V. M., 1927-8, *Geochem. Verteilungsgesetze der Elemente*, VII and VIII, Skr. Norske Videnskaps-Akad., Oslo, 1. Math-nat. Kl., 1926, No. 2, and 1927, No. 8 (see *Strukturber. Z. Kristallogr.*, 1928, **1**, 131).
 GROSS, S. T., 1941, *J. Amer. Chem. Soc.*, **63**, 1168.
 GROSS, S. T., 1943, *J. Amer. Chem. Soc.*, **65**, 1107.
 HALLA, F., and PAWLEK, F., 1927, *Z. phys. Chem.*, **128**, 49.
 HANAWALT, J. D., RINN, H. W., and FREVEL, L. K., 1938, *Ind. Eng. Chem., Anal. Ed.*, **10**, 457.
 HINTERBERGER, H., 1942, *Z. Phys.*, **119**, 1.
 JAMES, R. W., and WOOD, W. A., 1925, *Proc. Roy. Soc., A*, **109**, 598.
 KOLDERUP, N. H., 1924-5, *Bergens Museum Aarbok*, Nat. R., No. 2.
 LEE, E., and PARKER, R. C., 1946, *Nature, Lond.*, **158**, 518.
 LEHMANN, W. M., 1924, *Z. Kristallogr.*, **60**, 379.
 MOORE, W., and PAULING, L., 1941, *J. Amer. Chem. Soc.*, **63**, 1392.
 NELSON, J. B., and RILEY, D. P., 1945, *Proc. Phys. Soc.*, **48**, 85.
 VON OLSHAUSEN, S., 1925, *Z. Kristallogr.*, **61**, 463.
 PETERSEN, M., 1941, *J. Amer. Chem. Soc.*, **63**, 2617.
 RAMSDELL, L. S., 1925, *Amer. Mineral.*, **10**, 281.
 RENCKER, E., and BASSIÈRE, M., 1936, *C.R. Acad. Sci., Paris*, **202**, 765.
 RICHMOND, W. E., and WOLFF, C. W., 1938, *Amer. Mineral.*, **23**, 799.
 STARKIEWICZ, J., SOŚNOWSKI, L., and SIMPSON, O., 1946, *Nature, Lond.*, **158**, 28.
 STRAUMANIS, M., 1942, *Z. phys. Chem.*, **52**, 127.
 TRZEBIATOWSKI, W., 1937, *Roczn. Chem.*, **17**, 73.
 VLASAK, F., and TROUSIL, Z., 1934, *Věda přírodn.*, **15**, 164 (see *Strukturber.*, **3**, 258).
 WILMAN, H., 1940, *Proc. Phys. Soc.*, **52**, 323.
 VON ZEIPFEL, E., 1935, *Ark. Mat. Astr. Fys.*, **25** A, No. 8; *Chem. Zentr.*, **2**, 1326.

Solid Diagrams Illustrating Resonance Phenomena

By W. A. PROWSE

University of Durham

MS. received 17 April 1947, in amended form 15 July 1947

ABSTRACT. Three dimensional vector loci are used to express the properties of resonant circuits and of the tuned transformer. An indication is given of the application of the method to other problems, including a simple low frequency selector circuit.

§1. INTRODUCTION

THE study of the steady state behaviour of systems possessing stiffness, inertia and resistance forms a large part of any treatment of alternating current phenomena, of acoustics and of vibrating systems in general. It is not proposed to add further detail to this type of study, but rather to discuss a method of representing the behaviour of such systems which has been found attractive to students, which presents the solutions in a form which appeals to the imagination, and which is thought to afford a convenient visual summary of the phenomena concerned.

§2. SIMPLE RESONANT SYSTEMS

Consider a system whose behaviour is represented by the harmonic equation (figure 1)

$$L \frac{d^2 q}{dt^2} + R \frac{dq}{dt} + \frac{1}{C} q = V, \quad \dots\dots (1)$$

where q is the charge on the condenser and V the applied electromotive force.

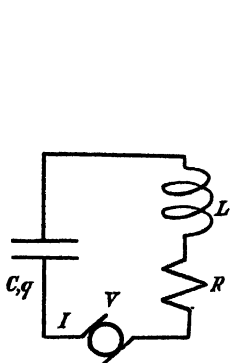


Figure 1.

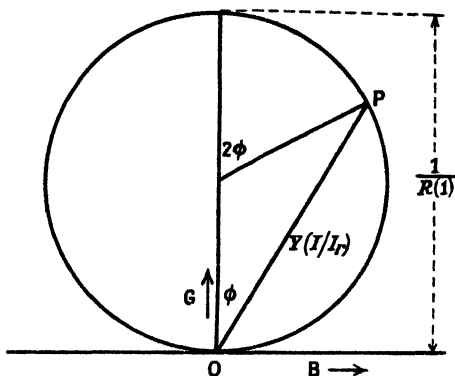


Figure 2. Vector locus of admittance (OP) of circuit of figure 1. Symbols in brackets denote alternative quantities. For the parallel circuit OI represents V/V_r .

The steady state solution may be written in terms of the impedance vector Z or of the admittance vector Y . Let these be divided into real and imaginary components, so that

$$Z = R + jX$$

and

$$\left. \begin{aligned} Y &= G + jB \\ &= (R - jX)/(R^2 + X^2) \end{aligned} \right\} \dots\dots (2)$$

For variations of X , the locus of Z is a straight line in the complex plane and that of Y is a circle of radius $1/2R$ (Hague 1945) (figure 2).

The latter result depends only on the relation (2) and not on the means whereby the variation of X is effected. If now a three-dimensional diagram be plotted, with the variable parameter responsible for changes in X along an axis at right angles to the G and B axes, a model of the type shown in figure 3 is obtained. It will be noted that the form of the curve is essentially a single circular turn on a straight line.

In this model the third variable is ω/ω_r , where ω is the angular frequency of V and ω_r the resonant value, $1/\sqrt{LC}$. By suitable illumination the three projections are obtained, showing respectively the variation of B and of G with frequency and the circle diagram, the latter, of course, being the vector locus of the admittance.

When used to illustrate current resonance in a series circuit, it is noted that the current I is related to the current I_r , at resonance, by

$$I/I_r = (1 - jX/R)/(1 + X^2/R^2),$$

so that the circle may be made of unit diameter. Here the chord from the point O (figure 2) to a given point on the circumference represents the vector value of the current, in terms of the current at resonance, for the frequency chosen. It is immediately apparent from the model that (i) the variations of conductance and of susceptance through resonance are simply two aspects of one process, (ii) the total variation of the susceptance is equal to the total variation of the conductance,

(iii) the maxima and minima of the susceptance curve occur on the steep parts of the conductance curve, (iv) the phase change in the neighbourhood of resonance is very rapid, (v) the phase at resonance is zero and (vi) the current is a small quantity in quadrature with the voltage at points well removed from the resonance peak.

§ 3. THE TUNED TRANSFORMER

Let the secondary circuit be of resistance R_2 and reactance X_2 and let it be coupled to the primary circuit by mutual inductance M . The increments of primary resistance ΔR and reactance ΔX , caused by the presence of the secondary are given by

$$\Delta R = \omega^2 M^2 \frac{R_2}{R_2^2 + X_2^2} \quad \dots\dots(3)$$

$$-\Delta X = \omega^2 M^2 \frac{X_2}{R_2^2 + X_2^2} \quad \dots\dots(4)$$

showing an exact correspondence in form with the real and imaginary components of Y (equation (2)).

The increase in impedance of the primary circuit can thus be represented by the same kind of solid diagram; in fact, the projections of the solid curve are perhaps most familiar as representations of the transferred impedance. Here the radius of the circle is $\omega^2 M^2 / 2R_2$, so that the representation is exact for tuning the circuit by varying C_2 or L_2 , but if the applied frequency causes the variation, the diagram is only accurate for sharply tuned circuits.

§ 4. MORE GENERAL METHODS

For purposes of calculation and for more general application (including parallel resonance when the tuning is reasonably sharp) the solution may be expressed in terms of Q and of the ratio of the impressed frequency ω to the resonant frequency ω_r (where ω_r is defined as the frequency for zero phase) (Lee and French 1943).

For the series circuit, since $Q = \omega_r L / R$ and $\omega_r = 1 / \sqrt{LC}$, we have $R = 1 / (Q\omega_r C)$. Substituting these values in equation (1) leads to

$$CV/q = (1 - \omega^2/\omega_r^2) + (j/Q)(\omega/\omega_r),$$

where V and q represent the vector values of the applied voltage and of the charge on the condenser respectively. q_r the charge at the resonant frequency, is given by

$$q_r/q = \omega/\omega_r - jQ(1 - \omega^2/\omega_r^2)$$

and

$$I/I_r = Y/Y_r = \frac{1}{1 - jQ(\omega_r/\omega - \omega/\omega_r)}. \quad \dots\dots(5)$$

In the complex plane (5) represents a unit circle in which $\tan \phi = Q(\omega_r/\omega - \omega/\omega_r)$ and, since the angle at the centre is 2ϕ the solid curve may be plotted by means of a piece of graph paper wrapped round a cylinder. This procedure, followed by shaping a piece of wire to fit the curve, was used in making the solid models illustrated.

§ 5. OTHER APPLICATIONS

The method has been found valuable for illustrating the properties of tuned transmission lines and of dielectrics containing charged particles, elastically

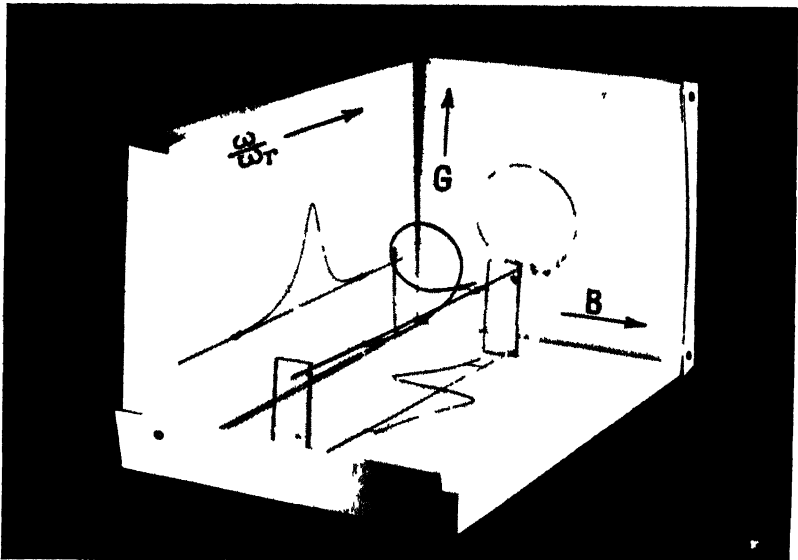


Figure 3.

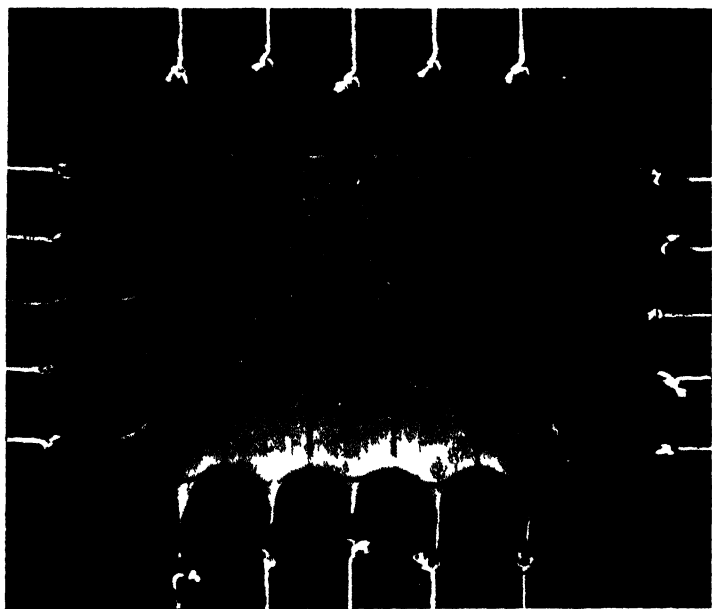


Figure 1 Rubber sheet stretched by factors of 2.5 and 2.0 in two perpendicular directions

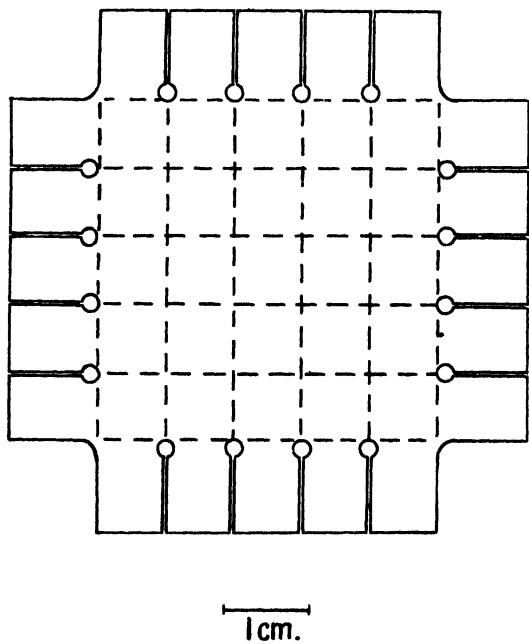


Figure 1 (a) 1 cm of unstretched sample

bound or aperiodic. In addition, a simple low frequency selector circuit has been developed. This, in effect, is an interstage coupling consisting of two transformers, one tuned slightly above and the other slightly below the required frequency, with the output voltages in opposition. It will be realized that, whereas the output of each transformer separately is represented by a line drawn from the origin to a point on the circumference of the circle (figures 2 and 3), the combined output is the chord joining the ends of two such lines. Thus the maximum output can be the same as that of either component separately, but it drops much more rapidly towards zero on either side of the peak. Satisfactory discrimination against the second harmonic is obtainable at 9 cycles per second.

ACKNOWLEDGMENT

I wish to take this opportunity of thanking the Research Committee of the Durham Colleges for their generous support, particularly in providing the oscillator used.

REFERENCES

- HAGUE, 1945, *A.C. Bridge Methods*, Chap. II (5th Ed., Pitman).
LEE, A., and FRENCH, W., 1943, *University Radio Conference Proceedings*, p. 67.

Stresses and Birefringence in Rubber subjected to General Homogeneous Strain

By L. R. G. TRELOAR

British Rubber Producers' Research Association

MS. received 23 January 1947

ABSTRACT. A method is described by which a sheet of rubber may be subjected to the most general type of homogeneous strain, with simultaneous measurement of the principal stresses and the birefringence. As predicted by the molecular network theory, the birefringence is found to be proportional (*a*) to the difference of the squares of the principal extension ratios, and (*b*) to the difference of the principal stresses in the sheet.

An examination of the stresses shows that these cannot be accurately represented by the equations derived from the network theory. A closer approximation is given by a more general theory of Mooney, the simplest form of which gives a satisfactory account of the behaviour of a swollen rubber. For a dry rubber, on the other hand, the Mooney formula, though better than the network theory, does not adequately represent the experimental data.

§ 1. INTRODUCTION

IN a previous paper (Treloar 1947 a) the author has dealt with the theory of the optical properties of strained rubber on the basis of a model in which the rubber was envisaged as a network of long-chain molecules of randomly-jointed links. These links were assumed to be optically anisotropic, having two different optical polarizabilities, in directions parallel and perpendicular to their length.

The theory was applied to the case of the most general homogeneous deformation of rubber, and it was shown that, for each of the three principal directions of strain, there is an associated optical polarizability, by which the refractive index for light having the corresponding direction of electric vector is determined. If λ_1 , λ_2 and λ_3 are the principal extension ratios, and n_1 , n_2 and n_3 the corresponding refractive indices, the double refraction for light propagated along the direction λ_3 may be represented by the equation

$$n_1 - n_2 = \frac{(\bar{n}^2 + 2)^2}{\bar{n}} \cdot \frac{2\pi}{45} N(\alpha_1 - \alpha_2) v_r^{1/3} (\lambda_1^2 - \lambda_2^2), \quad \dots\dots(1)$$

in which \bar{n} is the mean refractive index, N the number of molecular chains per c.c. and $\alpha_1 - \alpha_2$ the difference of polarizabilities for the link of the chain. The quantity v_r refers to the volume fraction of rubber in the case when the rubber is swollen in an optically inert solvent. Comparison of this result with the corresponding equation for the difference of principal stresses (Treloar 1947 a), i.e.

$$t_1 - t_2 = NkTv_r^{1/3}(\lambda_1^2 - \lambda_2^2), \quad \dots\dots(2)$$

where t_1 , t_2 , t_3 are the principal stresses, k is Boltzmann's constant and T the absolute temperature, leads to a linear relation between the birefringence and the stress-difference, which has already been examined experimentally for the case of a simple elongation in an earlier paper (Treloar 1947 b).

§ 2. EXPERIMENTAL METHOD

In the present paper a study of the stresses and birefringence in the most general type of homogeneous strain is reported. For this purpose a sheet of vulcanized rubber, compounded according to the formula given in the paper referred to (Treloar 1947 b), was stretched in two directions at right angles in such a way that the strain was homogeneous, whilst the two principal extension ratios λ_1 and λ_2 in the plane of the sheet could be varied independently by the application of two sets of forces. The arrangement used permitted the simultaneous measurement of the principal stresses and the double refraction corresponding to any state of strain.

It is not easy to devise a suitable method of applying two unequal strains to rubber in two perpendicular directions. The method here adopted, though not ideal, particularly in that the maximum attainable strain was limited, seemed to provide the simplest means of achieving the desired result. It makes use of the principle of the guard plate to eliminate the effect of the non-uniformity of the strain in the vicinity of the edges of the sheet. The shape of the test piece is shown in figure 1. It was in the form of a square, with five projecting lugs on each side, to which strings were tied for the application of the loads. The surface of the sheet was marked out with a series of lines forming a square lattice, so that the state of strain at all points could be observed. With the sheet placed horizontally, the three middle lugs on one side were loaded by means of three equal weights attached to strings passing over pulleys, whilst the strings attached to the two outermost lugs were secured to a rectangular frame, mounted horizontally. The five strings on the opposite side were connected to the opposite side of the frame. Similarly, the three middle lugs on an adjacent side carried three equal weights, whilst the two

outer lugs, as well as the five on the side opposite to them, were connected to the frame. The 14 strings connected to the frame were provided with tension adjusters, and these strings, as well as the pulleys, could be moved sideways independently, so that each set of strings could be kept parallel and the two sets maintained perpendicular to each other. The appearance of the stretched sheet when the tension on the outer lugs was suitably adjusted is shown in figure 1; in all cases the strain within the rectangle ABCD was sensibly uniform. Under these conditions it was assumed that the stresses acting on the sides of the rectangle ABCD were defined by the loads applied to the three central lugs, and that the non-uniform edge region, though requiring the use of a slightly different tension on the outer lugs*, would be without effect on the stress in this inner, uniformly-strained region.

For the optical measurement a beam of light from a polarizer set at 45° to the directions of principal strain was passed normally through the sheet, the birefringence being measured by means of a Babinet compensator.

The measurements were made at room temperature, which varied between 17 and 22°C .

§ 3. EXPERIMENTAL RESULTS

The data obtained are given in tables 1 and 2. Table 1 refers to the dry rubber, and table 2 to the same rubber swollen in medicinal paraffin to $v_r=0.525$. Readings were taken in the order given in the tables. Generally about 10 minutes elapsed between the application of a load and the taking of readings. The forces f_1 and f_2 are the loads in grams applied to each of the three middle lugs; the areas of cross-section of the rubber in the unstrained state (A_0), over which these forces act are given in the table. The figures for optical retardation are given in terms of the number of sodium wavelengths phase difference between the two rays, this being the quantity actually measured. To convert this to difference of refractive indices ($n_1 - n_2$) it is only necessary to divide by the thickness of the sheet and multiply by the wavelength. The extension ratio λ_3 , and the corresponding thickness in the strained state, were determined from the measured λ_1 and λ_2 , assuming no change of volume on straining.

In addition to the data represented in tables 1 and 2, figures were obtained for the case of simple elongation, i.e. with $f_2=0$, using a parallel strip of rubber cut from the same sheet. These additional data are included in the graphical presentation of the results.

The optical data are presented in figures 2 and 3, omitting points (near the origin) corresponding to the case when $f_1=f_2$. In the first case (figure 2) the birefringence is plotted against the difference of the squares of the corresponding extension ratios, $\lambda_1^2 - \lambda_2^2$. While the expected linear relationship (equation (1)) is approximately borne out, there are for the dry rubber, slight departures which are probably greater than the errors of measurement. For the swollen rubber, on the other hand, the agreement with the theoretical form is very close. In the second case (figure 3) the birefringence is plotted against $t_1 - t_2$, the difference of the principal stresses. Again, the proportionality is more exact for the swollen than for the dry rubber.

* It was found experimentally that the difference of tension on the outer strings was in fact very slight.

Table 1. Dry rubber. Thickness = 0.082 cm. $A_0 = 0.0648 \text{ cm}^2$

f_1 (gm.)	f_2 (gm.)	λ_1	λ_2	Retardation (sodium wavelengths)
100	100	1.07 ₀	1.08 ₃	- 0.02
100	200	0.97 ₈	1.29 ₇	- 0.55
200	200	1.16 ₇	1.20 ₉	- 0.07
300	200	1.46 ₀	1.09 ₁	+ 0.54
300	100	1.58 ₂	0.90 ₅	+ 1.06
300	300	1.35 ₀	1.37 ₀	- 0.05
300	400	1.18 ₃	1.88	- 0.80
200	400	0.93 ₃	2.04	- 1.50
100	400	0.78 ₁	2.14	- 2.00
400	400	1.60	1.72 ₅	- 0.13
500	400	2.24	1.51 ₃	+ 0.64
500	300	2.42	1.09 ₆	+ 1.39
500	200	2.51	0.87 ₅	+ 2.05
500	100	2.64	0.72 ₅	+ 2.80
500	500	2.07	2.10	- 0.05
600	500	2.68	1.87 ₅	+ 0.54
400	600	1.25 ₀	2.98	- 1.55
300	600	0.98 ₀	3.06	- 2.31
200	600	0.79 ₁	3.14	- 3.05
100	600	0.67 ₀	3.21	- 3.76
600	600	2.34	2.63	- 0.18

Table 2. Rubber swollen to $v_r = 0.525$ in paraffin. Thickness
(swollen) = 0.099 cm. $A_0 = 0.0985 \text{ cm}^2$

f_1 (gm.)	f_2 (gm.)	λ_1	λ_2	Retardation (sodium wavelengths)
100	100	1.08 ₈	1.08 ₈	- 0.02
100	200	0.99 ₆	1.30 ₅	0.40
200	200	1.21 ₅	1.19 ₃	+ 0.01
300	200	1.55	1.06 ₈	+ 0.51
300	100	1.66	0.88 ₂	+ 0.93
300	300	1.41	1.36 ₄	+ 0.03
100	400	0.80	2.04	- 1.46
200	400	0.95 ₆	1.95 ₃	- 1.02
300	400	1.24	1.83	- 0.51
400	400	1.65	1.65 ₄	0.00
400	500	1.50	2.12	- 0.46
500	500	1.97	1.94	+ 0.02
500	300	2.31	1.10	+ 1.06
500	200	2.43	0.85 ₈	+ 1.62
500	100	2.48	0.73 ₆	+ 2.07

Figure 4 represents the difference of the two principal stresses, plotted against $\lambda_1^2 - \lambda_2^2$, for both samples. As with the birefringence, the swollen rubber appears

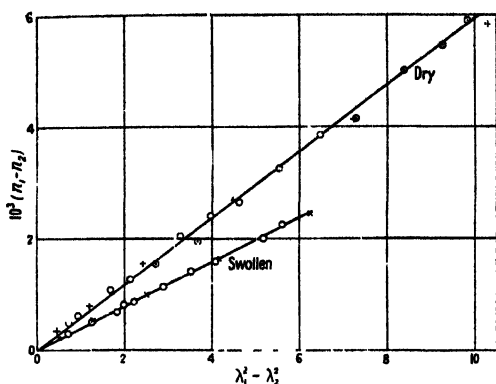


Figure 2. Birefringence plotted against difference of squares of principal extension ratios for dry and swollen rubber. The crosses refer to simple elongation.

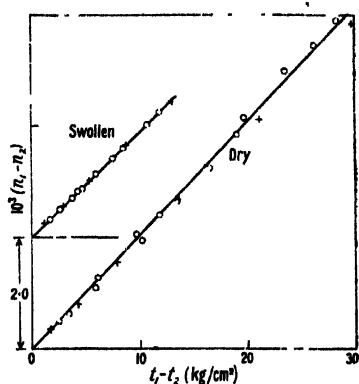


Figure 3. Birefringence plotted against difference of principal stresses for dry and swollen rubber. The crosses refer to simple elongation.

to agree with the theoretical form (2) within the experimental error, but the agreement in the case of the dry rubber is not exact.

§ 4. FURTHER EXAMINATION OF STRESSES

Whilst the evidence presented in figures 2, 3 and 4 suggests that the theoretical expectations regarding the dependence of the birefringence and of the principal stress difference on the two principal extensions in the plane of the sheet are approximately fulfilled, particularly in the case of the swollen rubber, an examination of each of the principal stresses separately reveals consistent deviations from the theoretical form, as will now be shown.

Since the stress t_3 normal to the sheet is zero, we should have, according to the theory previously outlined (Treloar 1947 a),

$$t_1 = G(\lambda_1^2 - \lambda_3^2), \quad \dots\dots (2a)$$

with a corresponding expression for t_2 . According to this equation, a plot of t_1 or t_2 , against $\lambda_1^2 - \lambda_3^2$ or $\lambda_2^2 - \lambda_3^2$ respectively, should yield a straight line. The actual results, shown in figures 5 and 6, present a more complex appearance for both the dry and swollen rubber.

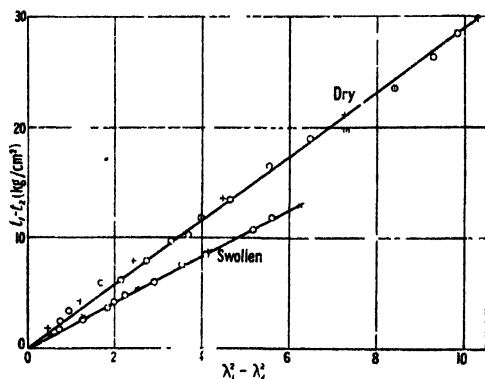


Figure 4. Difference of principal stresses plotted against difference of squares of principal extension ratios. The crosses refer to simple elongation.

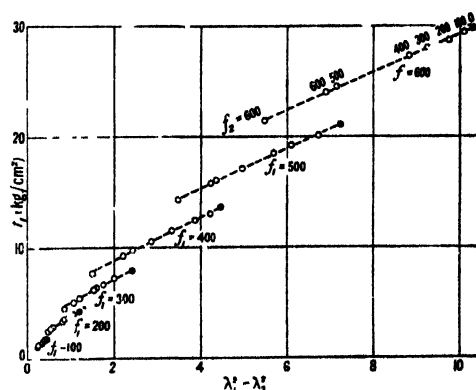


Figure 5. Dry rubber. Stress t_1 plotted against $\lambda_1^2 - \lambda_3^2$. The black circles refer to simple elongation. f_1 and f_2 are expressed in grams.

The method of plotting employed in these diagrams requires a little explanation. Firstly, it must be remembered that the stress (say t_1) is determined by the force (f_1) divided by the area of cross-section, which is itself a function of the strain, i.e. $t_1 = \lambda_1 f_1 / A_0$. Hence, while one of the forces is held constant, the stress corresponding to this force varies with changes in the second force, since such changes alter all the extension ratios. In order to simplify the presentation, the stress plotted is referred to as t_1 , and the corresponding force as f_1 . Thus two points appear on the diagram for each state of strain, one for each of the two principal stresses.

It is seen that the points fall on a set of lines, on any one of which f_1 is constant, while f_2 varies. This clearly means that t_1 is dependent on a more complicated function of the extension ratios λ_1 , λ_2 and λ_3 than the one represented by equation (2).

This important departure from the theoretical form exhibited when the principal stresses are plotted separately is rather surprising in view of the fair agreement obtained when only their difference is considered. Before discussing its significance, it seems desirable to examine very carefully the nature of the evidence, in order to make sure that it is not due to some spurious effect introduced by the experimental technique. Three possible sources of trouble will be discussed, namely (a) relaxation, (b) anisotropy in the original sheet, and (c) non-uniform strain round the edges.

(a) *Relaxation*

It is true that the attainment of equilibrium values of the stresses in strained rubber is a matter of considerable difficulty, requiring the breaking-down of temporary intermolecular "cross-links" by the use of higher temperatures or other means. No attempt was made to obtain genuine equilibrium values in the present experiments, and it is therefore necessary to consider whether the discontinuous array of figures 5 and 6 might arise from this cause.

There are two reasons for rejecting this hypothesis. Firstly, although it was observed that the values of the strain parameters obtained under the application of a given pair of forces depended to some extent on the order of application of the forces and on the previous history of the sample, nevertheless, the stresses, when plotted as in figure 5, invariably fall precisely on the appropriate discontinuous arrays. Secondly, it is clear that swelling should favour the approach to equilibrium, but the discontinuities are no less marked for a rubber swollen with nearly 100% of solvent than for the dry rubber.

(b) *Anisotropy in the original sheet*

It is conceivable that the departures from the theoretical form might arise from a difference of properties in different directions in the original sheet of rubber, introduced by the rolling process prior to vulcanization. Such an effect would arise from non-equilibrium or secondary linkages rather than from primary linkages, since the form of the stress-strain relations derived from the molecular theory is in no way dependent on the detailed structure of the molecular network (James and Guth 1943). If such anisotropy were present, it would be expected to reveal itself in a non-uniform swelling in different directions; but no such non-uniformity in swelling was observed. Furthermore, a latex rubber sheet,

prepared by a quite different process not involving rolling or pressing, gave a precisely similar type of diagram to that of figure 5. It is therefore concluded that an initial anisotropy cannot be responsible for the essential features of this diagram.

(c) *Non-uniformity of strain*

It has already been suggested that the guard-plate principle used in these experiments should eliminate any effect of non-uniformity of strain in the neighbourhood of the edges of the sheet.

In order to check this point, experiments were made by two independent methods. The first was by the use of parallel strips in simple elongation, using values of force per unit area of the unstrained section equal to the values of f_1 and f_2 in the tables. The points derived from these experiments, separately marked in figures 5 and 6, will be seen to fall exactly on the previously obtained arrays. The second check was obtained from measurements on a circular sheet clamped round its circumference and inflated. Details of the method have been described elsewhere (Treloar 1944 a, 1944 b). This corresponds to the special case where $\lambda_1 = \lambda_2$. The experiment was made on the actual sheet of swollen rubber used to obtain the data of figure 6. The results cannot conveniently be shown on this diagram, because the force could not simply be chosen to correspond to the values of f_1 and f_2 previously used. A plot of t_1 against $\lambda_1^2 - \lambda_2^2$, however, agreed to within 3% with the data in figure 6 for the case $f_1 = f_2$.

These independent experiments relating to the two extreme cases of simple elongation and 2-dimensional extension with $\lambda_1 = \lambda_2$ prove conclusively that the unexpected form of the results is not due to some defect in the technique of measurement. All the evidence suggests that the observed phenomenon represents a fundamental property of the rubber.

§ 5. THEORETICAL INTERPRETATION OF DATA

Mooney (1940) has discussed the form of the general stress-strain relations for a rubber on the basis of an assumed stress-strain relation in simple shear. For the particular case of a linear stress-strain relation in shear, he obtains, for the work of deformation

$$W_2 = C_1(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - 3) + C_2\left(\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} + \frac{1}{\lambda_3^2} - 3\right), \quad \dots \quad (3)$$

where C_1 and C_2 are physical constants of the material. The more general case of a non-linear shear relation may be resolved by adding further terms to equation (3). Thus the next term in the series would be of the type (Mooney, *loc. cit.* equation (40))

$$W_4 = A_4(\lambda_1^4 + \lambda_2^4 + \lambda_3^4 - 3) + B_4\left(\frac{1}{\lambda_1^4} + \frac{1}{\lambda_2^4} + \frac{1}{\lambda_3^4} - 3\right), \quad \dots \quad (4)$$

and the work of deformation would then be

$$W = W_2 + W_4.$$

From a rather more general standpoint, Rivlin (unpublished work) has argued that the work of deformation, or stored-energy function, must be a function of certain "strain invariants", which are themselves symmetrical functions of the

even powers of the extension ratios λ_1 , λ_2 and λ_3 . The simplest functions which may be chosen for the stored energy are

$$W_1 = G(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - 3), \quad \dots\dots(5)$$

and
$$W_2 = K\left(\frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2} + \frac{1}{\lambda_3^2} - 3\right). \quad \dots\dots(6)$$

Next in order of complexity are

$$W_3 = L(\lambda_1^4 + \lambda_2^4 + \lambda_3^4 - 3), \quad \dots\dots(7)$$

and
$$W_4 = M\left(\frac{1}{\lambda_1^4} + \frac{1}{\lambda_2^4} + \frac{1}{\lambda_3^4} - 3\right). \quad \dots\dots(8)$$

Rivlin's approach is perhaps simpler, and avoids the assumption that the form of the stress-strain relation for shear in an isotropic plane is independent of a stretch in the direction normal to that plane. However, the resultant stress-strain relations are the same whichever theory is adopted.

The stored-energy function derived from the molecular theory (Treloar 1947 a), from which equations (2) and (2a) are obtained, is

$$W = G(\lambda_1^2 + \lambda_2^2 + \lambda_3^2 - 3), \quad \dots\dots(9)$$

and is therefore equivalent to Mooney's equation (3) with the constant C_2 put equal to zero.

Let us now examine the stress-strain relations resulting from the first equation of Mooney (equation (3)). Corresponding to this stored-energy function, the principal stresses are given by equation (25) in his paper, which may be written in the form *

$$t_1 - t_3 = (G + K\lambda_2^2)(\lambda_1^2 - \lambda_3^2), \quad \dots\dots(10)$$

which, as will be seen, is a rather more complicated expression than equation (2) derived from the molecular theory, using the stored-energy function (9). In this expression G and K are physical constants expressible in terms of C_1 and C_2 in equation (3). In the case when the forces per unit area of the unstrained rubber, F_1 and F_2 are given, and $F_3 = 0$, we have the following equations relating the strains with the forces (or stresses)

$$\begin{aligned} t_1 &= \lambda_1 F_1 = (G + K\lambda_2^2)(\lambda_1^2 - \lambda_3^2), \\ t_2 &= \lambda_2 F_2 = (G + K\lambda_1^2)(\lambda_2^2 - \lambda_3^2). \end{aligned} \quad \dots\dots(10a)$$

These simultaneous equations enable the extension ratios to be determined when the applied forces F_1 and F_2 are given. Their solution has been obtained graphically, using the particular values $G = 1.0$, $K = 0.1$, for six values of F_1 combined in all possible ways with the same values of F_2 . This solution is represented in figure 7 in terms of the stress t_1 plotted against $\lambda_1^2 - \lambda_3^2$. Detailed comparison with figure 6 shows a very close quantitative correspondence between the theoretical and experimental points.

Another method of comparing the experimental data with the theoretical equation (10) is by plotting the experimental stress t_1 against $(1 + K\lambda_2^2/G)(\lambda_1^2 - \lambda_3^2)$,

* The constant G is not the same as Mooney's G .

choosing a suitable value of K/G . This type of plot is shown in figure 8, in which curve (b) relates to the swollen rubber. With $K/G=0.1$, the points fall very nearly on a straight line.

Thus Mooney's equation (3) gives a completely satisfactory representation of the behaviour of this particular rubber in the swollen state.

It is interesting to examine some further properties of Mooney's equation. Firstly, for a pure shear, we have one dimension (say λ_2) unchanged, and equation (10) reduces to

$$t_1 = (G + K)(\lambda_1^2 - \lambda_3^2).$$

In this case, therefore, the behaviour is indistinguishable from that given by the simple equation (2a). Secondly, for the difference of the two principal stresses, in the general homogeneous strain, we have

$$t_1 - t_2 = (G + K\lambda_3^2)(\lambda_1^2 - \lambda_2^2). \quad (10b)$$

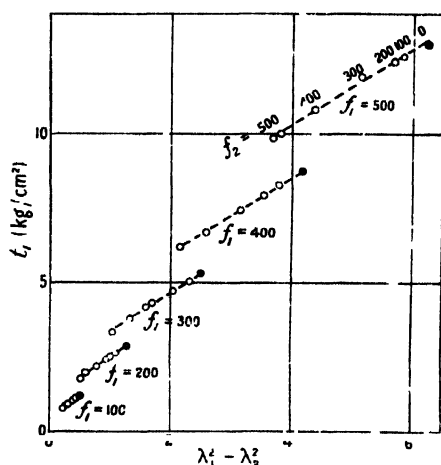


Figure 6. Swollen rubber. Stress t_1 plotted against $\lambda_1^2 - \lambda_3^2$. The black circles refer to simple elongation.

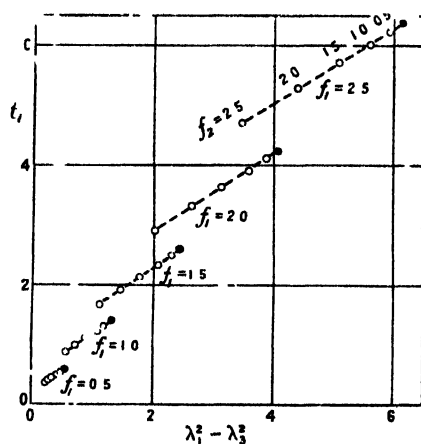


Figure 7. Stress calculated from Mooney's equation, using $G = 1.0$, $K = 0.1$.

If the forces are applied in the directions λ_1 and λ_2 , the extension ratio corresponding to the thickness direction is always less than unity, and λ_3^2 is generally quite small. If, therefore, K is also small compared with G , the term $K\lambda_3^2$ may be neglected, except at very small strains, i.e. equation (10b) reduces to the simpler form (2). This accounts for the observed linear relation between $t_1 - t_2$ and $\lambda_1^2 - \lambda_2^2$.

Turning now to the dry rubber, we find rather a more complicated picture. By a suitable choice of K/G it is possible to represent the stress t_1 as a continuous function of $(1 + K\lambda_3^2/G)(\lambda_1^2 - \lambda_3^2)$, but this function is not linear (figure 8 (a)). From this observation it is concluded that Mooney's equation represents a part, but not all of the deviation from the molecular theory (equation 2a). To account for the whole discrepancy it would probably be necessary to add further terms to the stored-energy function, such as those represented by equations (7) or (8). Since Mooney's equation (3) is based on a linear stress-strain relation in simple shear, it follows that the dry rubber is non-linear in shear. This fact may also

be deduced by direct interpolation from the experimental data. In other rubbers examined by the author, a non-linear shear relation has been obtained (Treloar 1944), but Mooney (1940) has reported a linear relation up to 200% shear.

§ 6. CONCLUSION

The conclusion to be drawn from the preceding analysis of the stress-strain relations may be summarized in the following way. As a first approximation the equations derived from the statistical treatment of a molecular model provide a basis for the interpretation of the elastic properties of rubber. A closer approximation is obtained by including an additional term in the stored-energy function, i.e. by the use of Mooney's equation (3). This appears to give an accurate representation of the properties of a swollen rubber, but is still inadequate when applied to a dry rubber, for which a third approximation, including at least one more general term, is likely to be required.

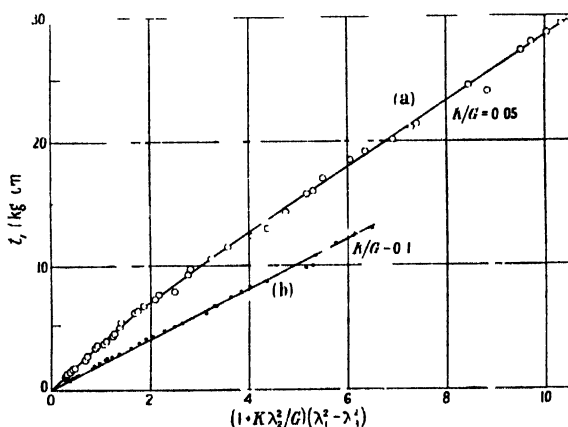


Figure 8. Data from figures 5 and 6 plotted with $(1 + K\lambda_2^2/G)(\lambda_1^2 - \lambda_2^2)$ as abscissae.
(a) Dry rubber, $K/G = 0.05$. (b) Swollen rubber, $K/G = 0.1$.

It is necessary to emphasize that the use of the second or higher approximations does not in itself throw any light on the physical mechanism responsible for the observed behaviour. It is, in fact, the 3-dimensional analogue of simple curve-fitting. It must also be borne in mind that any physical model that might be postulated must necessarily lead to results which may be represented by some combination of the stored-energy functions discussed in § 5.

ACKNOWLEDGMENTS

The author desires to express his thanks to his colleagues Dr. G. Gee and Mr. R. S. Rivlin, for helpful discussions in the course of this work, which forms part of the programme of fundamental research on rubber undertaken by the Board of the British Rubber Producers' Research Association.

REFERENCES

- JAMES and GUTH, 1943, *J. Chem. Phys.*, **11**, 455.
- MOONEY, 1940, *J. Appl. Phys.*, **11**, 582.
- RIVLIN. Unpublished work.
- TRELOAR, 1944 (a), *Inst. Rubber Industry Trans.*, **19**, 201.
- TRELOAR, 1944 (b), *Trans. Faraday Soc.*, **40**, 59.
- TRELOAR, 1947 (a), *Trans. Faraday Soc.*, **43**, 277.
- TRELOAR, 1947 (b), *Trans. Faraday Soc.*, **43**, 284.

The Dielectric Properties of Water and Heavy Water

BY C. H. COLLIE, J. B. HASTED AND D. M. RITSON

Clarendon Laboratory, Oxford

MS. received 26 March 1947

ABSTRACT. Measurements by a number of different methods of the dielectric constant and loss angle of water and heavy water at three widely separated wavelengths in the region of anomalous dispersion, $\lambda=10$ cm., 3 cm., 1.25 cm., are described. Results show that the Debye equations are exactly obeyed, there being a single time of relaxation varying with the temperature; the ratios of the relaxation times of water and heavy water are found to be in quantitative agreement with the ratios of the viscosity over a temperature range, suggesting that the re-orientation mechanism in these liquids is the same as that of viscosity. The value of the activation energy of the re-orientation process is discussed. The value of 5.5 found for the optical dielectric constant gives a reasonable result for the dielectric constant of water on the Onsager theory. This theory is discussed and shown to give a relaxation equation of the Debye type for alternating fields.

§ 1. INTRODUCTION

SINCE 1900 the anomalous dielectric dispersion of water at very short wavelengths has been widely investigated, and has been summarized by Dorsey (1940). The main qualitative features of the phenomenon are no longer in doubt, but quantitatively the results of different workers show wide divergencies. The significance of the problem lies in its theoretical bearing on the structure of water, which is not only of fundamental importance but also of technical and biological interest. In addition, a knowledge of the dielectric properties of water is needed in calculations on the transmission of microwaves.

It therefore seemed important to apply the recently developed high-frequency techniques to obtain an accurate measurement of the refractive index n and absorption coefficient κ of water. It was also felt that parallel measurements on heavy water would provide comparative data of theoretical value. In this paper measurements are described at three widely separated wavelengths, namely 10.0 cm., 3 cm., and 1.25 cm., in the temperature range 0° – 75° C.

The theory of an ideal polar dielectric in an alternating field was first given by Debye in terms of a single relaxation time. It can be shown that the relation below holds generally for linear dielectrics, as is discussed later in this paper:

$$\epsilon - \epsilon_0 = \frac{\epsilon_s - \epsilon_0}{1 + j\omega\tau}, \quad \dots\dots (1)$$

in which τ is a time of relaxation, and ϵ is the complex dielectric constant at the angular frequency ω ; ϵ_0 is that part of the dielectric constant which is due to the atomic and electronic polarization, and is assumed to be real and independent of ω . The static dielectric constant is ϵ_s . This expression may be expressed in terms of the optical constants as follows:

$$\begin{aligned} \kappa^2 = \epsilon' &= \frac{\epsilon_s - \epsilon_0}{1 + \omega^2\tau^2} + \epsilon_0 = \frac{\epsilon_s - \epsilon_0}{1 + [\lambda_s/\lambda]^2} + \epsilon_0, \\ 2n\kappa = \epsilon'' &= \frac{(\epsilon_s - \epsilon_0)\omega\tau}{1 + \omega^2\tau^2} = \frac{(\epsilon_s - \epsilon_0)\lambda_s/\lambda}{1 + [\lambda_s/\lambda]^2}, \end{aligned}$$

where n is the real part of the refractive index, κ the absorption coefficient, ϵ' and ϵ'' the real and imaginary parts of the complex ϵ , λ the wavelength corresponding to a frequency ω , and λ_s the wavelength which corresponds to frequency $1/\tau$ (the so-called "Sprungwellenlänge").

These equations represent the fall of the dielectric constant from its static value ϵ_s to its optical value ϵ_ω , the fall being accompanied by a single broad absorption band in the neighbourhood of the characteristic wavelength λ_s . The present measurements were made with the object of finding out whether the simple expression with a single relaxation time does adequately represent the change in optical constants with frequency at a fixed temperature. As will be seen, this is in fact the case, so that one can also investigate:

- (1) the variation of λ_s with temperature. This can then be compared with other physical properties, such as the viscosity, whose temperature dependence is known.
- (2) the ratio of λ_s for water and heavy water.

§2 METHODS OF MEASUREMENT

The optical constants vary so rapidly with wavelength and temperature that no universal method is available.

The following methods, some of which have been described in previous publications, have been used:—

- (1) The change of the resonant frequency and damping of an E_{010} resonator, when an axial capillary tube filled with water is inserted, is measured (Willis Jackson 1946, Collie, Ritson and Hasted 1946).
- (2) A similar measurement using an H_{01} resonator (Penrose 1946, Collie Hasted and Ritson 1948).
- (3) Measurements of the absorption coefficient are made by direct transmission in a waveguide filled with the liquid under examination (figure 1, B). If two measurements are made of the absorption coefficient, one, κ_1 , in a rectangular guide of width a_1 , and a second, κ_2 , in a guide of width a_2 near cut-off, then it is easily shown (loc. cit.) that

$$\epsilon' = \frac{(c^2/\omega^2)(k_1^2\kappa_1^2 - k_2^2\kappa_2^2)}{\kappa_1^2 - \kappa_2^2} - (\kappa_1^2 + \kappa_2^2),$$

$$\epsilon'' = 2\kappa_1[\epsilon' - k_1^2 c^2/\omega^2 + \kappa_1^2]^{1/2},$$

in which $k_1 = \pi/a_1$ and $k_2 = \pi/a_2$ for the H_{10} wave in a rectangular waveguide.

This method affords a satisfactory way of measuring both n and κ . As can be seen from the form of the equation the method is a practicable adaptation of the ideal method of determining κ by determining the attenuation in free space, and n by measuring the critical wavelengths of propagation down a guide. The quantities which have to be measured accurately are the width of the guide and the wavelength. These are the two quantities which can be measured with the greatest precision, and the latitude in the choice of a_1 and a_2 is sufficient to allow the two measurements of the attenuation in the guide to be approximately independent measurements of n and κ the optical constants in free space. This method is not limited to rectangular waveguides, and some of the later measurements

have with advantage been made using an H_{11} wave in cylindrical waveguides of radii r_1 and r_2 , in which case $k_1 = 1.841/r_1$, and $k_2 = 1.841/r_2$.

- (4) Direct measurements of the absorption in a sufficiently wide waveguide yield κ directly (figure 1 A).

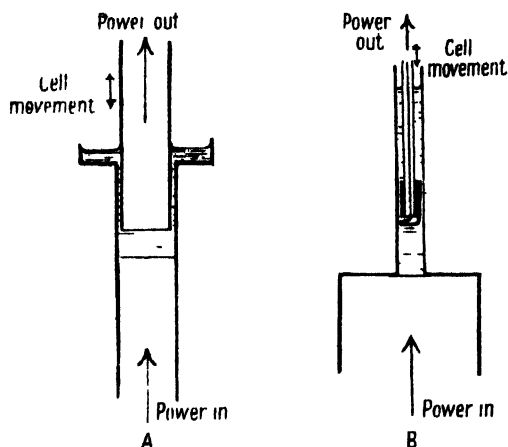


Figure 1. Principle of absorption measurement at 1 cm. and 10 cm.

The methods which have been used are shown in table 1.

Table 1. Methods of determination of n and κ

Wavelength (cm.)	H_2O		D_2O	
	n	κ	n	κ
10	E_{010} resonator. Absorption in two rectangular guides. Absorption in two circular guides.	Absorption in two rectangular guides. Absorption in two circular guides.	Absorption in two rectangular guides.	Absorption in two rectangular guides.
3.2	Absorption in two circular guides.	Absorption in wide rectangular guide. Absorption in two circular guides.	Absorption in two circular guides.	Absorption in two circular guides.
1.25	H_{01} resonator	Absorption in wide rectangular guide.	H_{01} resonator.	Absorption in wide rectangular guide.

§ 3. MEASUREMENT OF n AND κ AT 10 CM.

These measurements are relatively easy to make using radar receiving equipment whose main features were standardized during the war.

Power from a well screened reflex klystron (type CV 35) is led through a coaxial lead to a coaxial socket of the type first used by the General Electric Company in their research laboratories at Wembley. Commercial pyrotenax cable (type 147/1) is very suitable for the coaxial transmission line since the continuous outer coating facilitates the very high degree of screening needed if the fullest use is to be made of the high sensitivity obtainable (10^{-11} watt) with a superheterodyne receiver.

From the coaxial plug the power reaches the receiver, either through the absorption cell or through an II_{11} piston attenuator.

The construction of these absorption cells is shown diagrammatically in figure 1 B and the absorption cell using circular guides is shown in detail in figure 2.

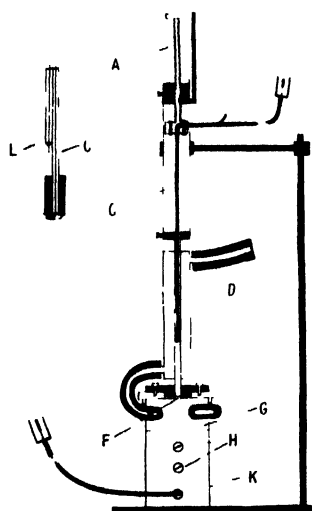


Figure 2
10-cm. absorption cell.

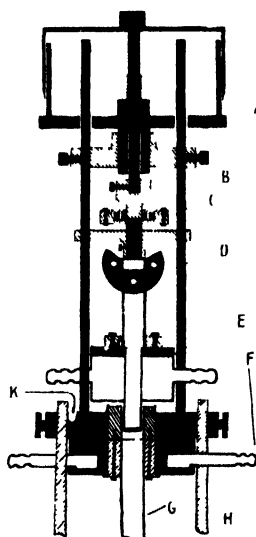


Figure 3
1-cm. absorption cell.

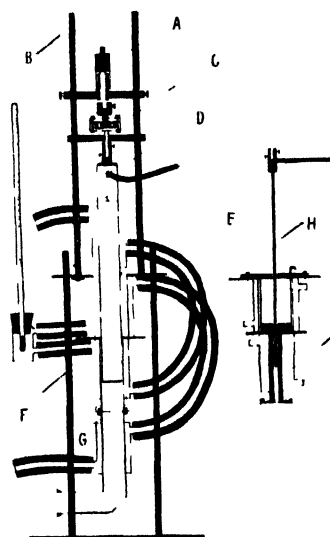


Figure 4
3-cm. absorption cell.

Power from the coaxial line E is fed into the rectangular wave guide K by means of a probe; a match is obtained with two stubs 3.75 cm. apart at H. A small fixed loop in this section enables the power level to be monitored on a crystal detector. From the top of the guide the power passes through a $\frac{1}{2}$ mm. mica window into a circular guide filled with water. The window is held in position by a plate secured by three countersunk screws. The joint is rendered watertight with Dow-Corning silicone stop-cock grease which provides an effective seal even at 100°C .

The interchangeable circular guide is made from copper tube straightened on a silver steel mandrel. The two diameters used were 1.261 cm. and 0.576 cm. As far as could be judged by measurement of the open ends and the fit of a steel plug the guides were uniform to 1 part in 200, and this is confirmed by the absence of systematic variations of the measured absorption coefficient.

After transmission through the water the power is again picked up by a loop at the end of a movable pyrotenax coaxial cable C. The loop can be withdrawn and its position read by means of a 12 cm. screw movement A.

The loop is held central in the guide by means of a loosely fitting plunger (see inset). The loop extends about $1\frac{1}{4}$ mm. into the liquid and is nearly the same diameter as the tube C. Spacers can be inserted into the movement to allow different portions of the wave guide to be explored by the loop.

A subsidiary electrode L protruding from the cable C enables the conductivity of the water to be measured during the course of measurement. The absolute conductivity is obtained by calibrating the apparatus with N/50 KCl. The true value of ϵ'' is obtained from its apparent value ϵ''_a by the relationship $\epsilon''_a = \epsilon'' + 4\pi\sigma/\omega$ in which σ is the D.C. conductivity in E.S.U. If conductivity water is used and the joints are made so that as little solder as possible is in contact with the water the conductivity correction can be kept below 2% of ϵ'' at 80°C. The magnesium oxide filling of the pyrotenax must be prevented from coming into contact with the water by a well baked layer of shellac varnish.

For measurements at 3.2 and 1.25 cm. this correction is sufficiently small to justify the omission of a continuous measurement of the conductivity. This was, however, always measured at the end of a run to verify that excessive corrosion had not occurred.

At all frequencies freshly boiled water must be used to prevent the formation of bubbles. For the same reason the temperature should be progressively lowered during a series of measurements. The heavy water used throughout was of 99.5% purity.

All the relevant parts of the apparatus were water jacketed, and water from a thermostat was circulated through the jackets through lagged pipes.

The temperature was measured in the water stream by a thermometer which had been compared with a set of thermometers by Anschütz, which had themselves been compared with thermometers carrying an N.P.L. certificate.

The measurement of the absorption coefficients is made in two parts. First the receiver responses R_1, R_2, \dots, R_n are measured for a series of positions of the loop in the absorption cell. The piston attenuator is then substituted for the cell and a measurement made of the position of the attenuator loop at which the receiver responses are the same as one of the previously noted values R_1, R_2, \dots, R_n . To verify that the power output of the transmitter and the sensitivity of the receiver have not changed during the measurement, the response R_1 is checked after every reading of R_n . In this way it is possible to plot a curve whose ordinate is the displacement of the loop in the absorption cell and whose abscissa is the displacement of the loop in the H_{11} attenuator which produces the same change in receiver response. This curve should be a straight line whether the receiver response is linear or not.

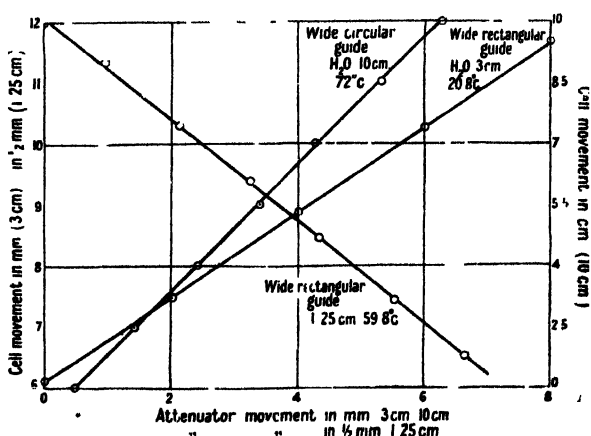


Figure 5. Typical absorption plots.

This offers a valuable check and enables errors in the experimental arrangements to be detected and eliminated. The two most obvious sources of error are pick-up and the production of unwanted modes in the absorption cell. Since the former is most troublesome at low signal strengths and the latter when the path in the water is small and the signal consequently strong, it is very improbable that these sources of error will compensate each other. Typical straight lines are shown in figure 5. The slope of these lines was evaluated by the method of least squares and the absorption coefficient calculated from the theoretical attenuation of the attenuator.

§4 THE RECEIVER

An Admiralty receiver type P 50 was used in these measurements. The band width of 3 Mc/s. was sufficient to embrace the small frequency wanderings of the transmitter and local oscillator klystrons. Although more complicated than a simple crystal rectifier and sensitive galvanometer it has so many advantages that its use is almost essential if an overall accuracy of 1% is aimed at.

The main advantages are:—

- (a) A constancy of calibration lasting for several hours, and in favourable circumstances over a period of days.
- (b) A factor of at least 10^4 in the usable sensitivity.

The increased sensitivity is of great importance in the type of work being described because:—

- (i) It enables all coaxial leads to be of sufficient length (about 2 m. of pyrotenax) to damp out the standing waves due to inevitable mismatches at plugs and junctions.
- (ii) It enables power levels to be measured with a piston attenuator; otherwise the use of these admirable instruments is precluded by their large insertion loss.

The use of an attenuator allows the receiver to be used as a monitor so that it is sufficient to read the second detector current on a 0–100 microammeter. A suitable resistance net work allows a 40 db. range to be covered without changing the gain of the I.F. stages.

§5. MEASUREMENTS AT 1.274 CM.

Accurate measurements at this wavelength offer some difficulty since the degree of frequency stability required is relatively much greater if an I.F. amplifier of 3 Mc/s. bandwidth is to be used, while the valves are inherently less frequency stable than klystrons operating on a longer wavelength, owing to the very small clearances and high power dissipation involved. Satisfactory measurements were made possible by modulating the frequency of the output klystron by means of a 10–40 volt saw-tooth voltage of recurrence 500 c.p.s. applied to the reflector. The output of the receiver is then to a first approximation independent of the absolute frequency, since the I.F. frequency now sweeps through the whole effective spectrum of the I.F. receiver. Satisfactory results were obtained provided outside conditions were kept as constant as possible. This was done by enclosing both the output and local oscillator klystrons in copper screening boxes cooled by a

constant stream of compressed air; in addition the voltage control was improved by replacing the neon stabilizers in the power pack by dry batteries. Some attempts were made to use the "two waveguide" method which had proved so convenient at 10 cm. It was not very satisfactory owing to the difficulty of matching into and out of a waveguide only 0.9 mm. in diameter, and the method was abandoned. The refractive index n was measured by the resonator method previously described (Collie, Hasted and Ritson 1948), and κ was determined by direct measurement of the attenuation in the absorption cell, shown in figure 3, and diagrammatically in figure 1 B. The water is contained in the upper part of the vertical rectangular guide G of internal cross section 10.75 mm. \times 4.5 mm., supported on a thin (0.15 mm.) mica window firmly clamped and greased in position. After traversing the water the wave is picked up by a movable guide (internal cross section 8.75 mm. \times 3.5 mm.) fitted with a distrene or rubber window about 2 mm. thick. This type of pick-up is very satisfactory and is much easier to match into the water than the loop and coaxial line used at longer wavelengths. The wave is led to the mixer by the right-angle bend at D and a length of flexible waveguide. A pair of stubs provides sufficient matching. In view of the rather open design of the cell it was thought worth while to measure the difference in temperature between the water in the cell and the water circulating in the water jackets. Measurement with a thermocouple showed that after ten minutes at any fixed temperature the difference in temperature between the water in the cell and the thermometer chamber was less than 0.1°C. The attenuation produced by a known thickness of water was measured as before by direct comparison with a piston attenuator.

The greatest source of error in absorption measurements of this sort lies in the possibility that new modes are created when the wave passes through the window into the water-filled guide. Each mode will be differently attenuated so that little precise information can be obtained from the measured attenuation curves, which are not exponential. In these experiments the production of unwanted modes was minimized by using a thin window whose flatness could be investigated by reflected light after it had been mounted in the cell. Distortion due to differential expansion was avoided by greasing, but not sticking, the window to the copper flange. A watertight joint up to 80° was ensured by using non-melting silicone grease. The freedom from unwanted modes, and overall accuracy of measurement, is shown by the achievement of straight-line plots for cell movement against piston attenuator movement. Typical plots at all wavelengths are shown in figure 5.

At this wavelength good straight lines through more than 3 mm. of water were obtained with this design of cell. There is, however, disagreement between these results and those we reported in an earlier paper, which is due to two facts:—

- (i) The design of the earlier cell did not permit the propagation of pure modes, so that curved logarithmic absorption plots were obtained; this curvature was erroneously described as geometrical in origin.
- (ii) In order to eliminate this effect measurements were taken with a large dead space of liquid, but through small thicknesses, with very high receiver sensitivity. Repetition showed that these measurements were vitiated by pick-up.

In the revised method both these sources of error have been eliminated beyond reasonable doubt, and the new results are taken as accurate to $\pm 1\%$.

§ 6. MEASUREMENTS AT 3.213 CM.

Measurement of κ at this wavelength was carried out with a rectangular wave guide cell similar to that described for use at 1.274 cm. It is shown in figure 4.

A type CV 87 reflex klystron was used as the source of power; this valve has a coaxial line output and is connected to a waveguide system by a probe. After passing through a unidirectional feed and a flap attenuator the power is matched by stubs through a thin mica window G into the water-filled cell. The small amount of power (about 10^{-6} W) led off by the one way feed is coupled into a high-Q resonator, the output of which is fed to a crystal rectifier and a high sensitivity galvanometer. In this way both the power level and the frequency stability of the wave entering the cell are continuously monitored throughout a series of readings.

The method of measurement was similar to that used at 1.274 cm. The cell was so designed that a number of fine cylindrical waveguides could be substituted for the rectangular guide, and a concentric line and pick-up loop for the output guide. These parts are shown in figure 4, and when they are in use the apparatus is similar to that used at 10 cm. for the measurement of both n and κ . The diameters of tube used were 2.51 mm. and 3.77 mm. As can be seen from figure 7, excellent agreement was obtained between values of κ obtained by the two methods, and the fine cylindrical wave guide has been used throughout for κ measurements on heavy water, since less liquid is required. The cylindrical tubes were used throughout for measurement of n .

§ 7. MEASUREMENT OF WAVELENGTH AND ATTENUATION

The wavelengths used were 1.274 cm., 3.213 cm., and 10.00 cm. respectively. The two former were measured with a simple absorption wave-meter using an H_{11} mode. The absolute accuracy depends upon the diameter of the reamed hole. They are uncorrected for the dielectric constant of air, and probably accurate to better than $\frac{1}{2}\%$. The 10 cm. wavelength was measured with a high precision resonant cavity wavemeter (Admiralty type G 93), which could be calibrated against a crystal-controlled oscillator.

Three inductive piston attenuators using an H_{11} mode were used in these measurements. Their attenuation constants are given below:—

Wavelength (cm.)	Attenuator, diameter (mm.)	A
1.274	4.00	1.575
3.213	9.525	1.70
10.00	12.74	4.48

The attenuation constant κ is calculated from the relationship

$$\kappa = \frac{\text{displacement of the attenuator}}{\text{displacement through the water}} \times A.$$

The attenuation constant A is calculated from the relationship

$$-4\pi^2 A^2 / \lambda^2 + (1841)^2 / r^2 = \omega^2 / c^2.$$

An account of the theory and construction of these instruments has been given *inter alia* by Clayton, Houldin, Lamont and Willshaw (1946).

§ 8. RESULTS

Values of n and κ have been obtained at a sufficiently large number of temperatures to enable smooth curves to be drawn through the points and values at ten degree intervals from 0° to 75° C. for water and from 5° to 60° C. for heavy water to be given at each frequency. These values, which are estimated as accurate to $\pm 1\%$ (10:1 probability) and $\pm 0.4^\circ$ C. in temperature, are given in tables 2 and 3, together with the values calculated from a single value of λ_s which is also tabulated at each temperature. Some measured values at 3 cm. and the smooth curves through them are shown in figure 7. It will be seen that except for a very few values there is

Table 2. Dielectric properties of water

T (°C)	ϵ_s	λ_s	λ 1.27 cm		λ 3.21 cm.		λ 10.0 cm.	
			n	κ	n	κ	n	κ
0	88.2	3.34	4.94	2.86	7.28	2.86	9.03	1.37 exp.
			4.92	2.81	7.31	2.83	9.04	1.37 calc.
10	84.2	2.39	5.55	2.91	7.73	2.43	8.89	0.982 exp.
			5.60	2.92	7.88	2.39	9.00	0.989 calc.
20	80.36	1.80	6.25	2.86	8.08	1.97	8.83	0.739 exp.
			6.20	2.83	8.16	1.96	8.87	0.737 calc.
30	76.7	1.39	6.75	2.67	8.11	1.57	8.78	0.558 exp.
			6.71	2.64	8.25	1.58	8.70	0.559 calc.
40	73.1	1.12	7.0	2.40	8.20	1.29	8.53	0.442 exp.
			7.02	2.39	8.21	1.28	8.51	0.440 calc.
50	69.8	0.912	7.26	2.11	8.15	1.05	8.28	0.350 exp.
			7.23	2.10	8.12	1.04	8.33	0.348 calc.
60	66.6	0.760	7.30	1.87	7.99	0.865	8.09	0.281 exp.
			7.34	1.83	8.01	0.856	8.14	0.283 calc.
75	62.1	0.608	7.35	1.52	7.82	0.672	7.78	0.212 exp.
			7.34	1.50	7.80	0.663	7.87	0.218 calc.

- (ii) Physically attractive but mathematically invalid approximations. The so called "wire wave" methods are open to such approximations as can be seen from the rigorous treatment which has been given by Slevogt.
- (iii) The relationship

$$R = \frac{n^2 + \kappa^2 + 1 - 2n}{n^2 + \kappa^2 + 1 + 2n}$$

between the reflection coefficient R and the optical constants, shows R to be insensitive to changes in n and methods based on the measurement of R are correspondingly uncertain.

- (iv) Owing to lack of power, experimenters using the optical or free wave method have sometimes had to make measurements before the region of steady exponential attenuation had been reached.

The authors hope that they have avoided or overcome most of these difficulties. As has been already mentioned all power measurements have been made with a piston attenuator, so that the receiver acts only as a monitor. Where possible independent methods have been used to cross-check some of the results.

The most extensive series of independent measurements was made at $\lambda = 10$ cm. at which wavelength the E_{010} resonator method was compared with both versions of the two-waveguide method over a large temperature range. The results given in table 5 show satisfactory agreement. At $\lambda = 3.2$ cm. κ was measured using both cylindrical and rectangular guides. Reference to figure 7 shows that the two methods agree. Up to the present no method of cross-checking the 1.25 cm. results has been possible apart from doubtful tests of self-consistency and consistency with the results at other wavelengths.

Table 5. 10 cm. results showing agreement between various methods

$T(^{\circ}\text{C.})$	Resonator	n		κ	
		Rectangular waveguide	Cylindrical waveguide	Rectangular waveguide	Cylindrical waveguide
0	8.99 ± 0.09	9.05 ± 0.05	9.03 ± 0.05	1.37 ± 0.01	1.37 ± 0.01
10	—	8.89	8.89	0.98	0.982
20	8.86	8.83	8.83	0.74	0.739
30	8.68	8.78	8.78	0.56	0.558
40	8.50	8.53	8.53	0.435	0.442
50	8.28	—	8.28	0.345	0.350
60	8.09	—	8.09	0.28	0.281
70	7.87	—	7.78	0.23	—
80	7.71	—	—	0.20	—
90	7.55	—	—	0.17	—
100	7.38	—	—	0.15	—

§ 10. THEORETICAL

It is significant, though not unexpected, that water and heavy water, in common with many other chemically pure liquids, obey the Debye equations exactly and have a single time of relaxation. From these data the values of dielectric constant and loss angle can be calculated for any frequency. For completion of the work, measurements at wavelengths of 0.5 cm. and 1 mm. would be desirable, but technique is not yet advanced sufficiently for these wavelengths to be available.

By such measurements, greater accuracy than ± 1 in the value of ϵ_0 would be obtained.

It is also significant that the single time of relaxation should hold at all temperatures, indicating that the change between the various structures suggested by Bernal and Fowler (1933) is continuous. It would be difficult to imagine dipoles in two different crystalline structures simultaneously present at any given temperature taking the same time of relaxation. It is apparent that the absorption lines shown by water vapour in this region have no effect on the dielectric constant of the liquid.

The expression $\epsilon - \epsilon_0 = (\epsilon_s - \epsilon_0)/(1 + j\omega\tau)$ which reproduces the experimental results was originally obtained by considering a dielectric in which the internal field was a linear function of the polarization. It is well known that this assumption over-estimates the interaction between the dipoles so that the dielectric model becomes spontaneously polarized except at very high temperatures. This difficulty is considerably enhanced by increasing the value of ϵ_0 . Moreover, there is a discrepancy between Debye's τ and the experimental τ which arises from the assumption of the Lorentz internal field (see *Polar Molecules*, p. 92). It is therefore important to consider how the dielectric constant would depend upon frequency on theories such as those of Onsager or Bottcher (1942), which avoid these difficulties by making the interaction between the dipoles depend on the dielectric constant. The following treatment of Onsager's theory follows closely that of Fowler and Guggenheim (*Statistical Thermodynamics*, 1939) except that an alternating field and a relaxation time are introduced.

Consider first a molecule radius a , with a polarizability γ , reduced to a refractive index n by the Lorentz-Lorenz formula as follows:—

$$\gamma = a^3(n^2 - 1)/(n^2 + 2). \quad \dots\dots(1)$$

If the electric moment of the molecule is μ , and it is immersed in a continuous medium of dielectric constant ϵ_s , then by Onsager's theory there will be a "reaction" field R given by

$$\mathbf{R} = \mu \cdot 2(\epsilon_s - 1)/(2\epsilon_s + 1)a^3. \quad \dots\dots(2)$$

Where by definition $\boldsymbol{\mu} = \mu_0\mathbf{u} + \gamma\mathbf{R}$ and \mathbf{u} is the unit vector in the direction of the moment. Therefore

$$\boldsymbol{\mu} = \mu_0\mathbf{u} + \gamma\boldsymbol{\mu} \cdot 2(\epsilon_s - 1)/(2\epsilon_s + 1)a^3. \quad \dots\dots(3)$$

Combining (3) with (1) and re-arranging,

$$\boldsymbol{\mu} = \mu_0\mathbf{u}(n^2 + 2)(2\epsilon_s + 1)/3(2\epsilon_s + n^2), \quad \dots\dots(4)$$

and combining (4) with (2)

$$\mathbf{R} = 2(n^2 + 2)(\epsilon_s - 1)\mu_0\mathbf{u}/3(2\epsilon_s + n^2)a^3. \quad \dots\dots(5)$$

Now consider the application of an alternating field \mathbf{E} of frequency ω , and let the dielectric constant corresponding to this frequency be ϵ . Let the additional field this causes at the cavity be \mathbf{E}' .

Then by the Onsager theory

$$\mathbf{E}' = \frac{\mathbf{E} \cdot 3\epsilon}{2\epsilon + 1} + \frac{\mathbf{E}' \cdot \gamma \cdot 2(\epsilon - 1)}{(2\epsilon + 1)a^3}, \quad \dots\dots(6)$$

and by a similar argument as was used for equation (4)

$$\mathbf{E}' = \mathbf{E}(n^2 + 2)\epsilon/(2\epsilon + n^2). \quad \dots\dots(7)$$

The total electric force \mathbf{F} acting on the molecule is given by

$$\mathbf{F} = \mathbf{E}' + \mathbf{R}, \quad \dots\dots(8)$$

so that

$$\mathbf{F} = \frac{\mathbf{E}(n^2 + 2)\epsilon}{2\epsilon + n^2} + \frac{2(n^2 + 2)(\epsilon_s - 1)}{3(2\epsilon_s + n^2)a^3} \mu_0 \mathbf{u}. \quad \dots\dots(9)$$

One may now follow the conventional treatment as given, for example, in Fowler and Guggenheim. In evaluating the mean polarization, the effect of the relaxation time can be introduced as in Debye's original treatment by writing $\mu_0/3kT(1 + j\omega\tau)$ instead of the expression $\mu_0/3kT$ appropriate to the static case. With this modification one obtains after some reduction the result that

$$\epsilon = \frac{A}{1 + j\omega\tau} \cdot \epsilon_s + \frac{1}{2}$$

in which A is written for the expression

$$\frac{(n^2 + 2)^2 \cdot (2\epsilon_s + 1)}{(2\epsilon_s + n^2) \cdot 6kT} \cdot \mu_0^2.$$

Since $1/2\epsilon$ is small compared with unity no great error will be involved by substituting

$$\epsilon = \epsilon_s / (1 + j\omega\tau).$$

Thus

$$\epsilon = n^2 + A / (1 + j\omega\tau) - A / 2\epsilon_s.$$

As $\omega \rightarrow \infty$, $\epsilon_s - (n^2 - A / 2\epsilon_s) \rightarrow A$. If ϵ_0 is defined so that $A = \epsilon_s - \epsilon_0$, then $n^2 - \frac{1}{2} = \epsilon_0$. We have, therefore, $\epsilon - \epsilon_0 = (\epsilon_s - \epsilon_0) / (1 + j\omega\tau)$, with $\epsilon_0 + \frac{1}{2} = n^2$. The effective deviation is of the order of $1/4\epsilon^2$ and will therefore be vanishingly small.

We can therefore say that it is impossible to differentiate experimentally between a linear internal field and an Onsager field.

The above calculation also shows that the dielectric constant differs from the value calculated from the relaxation time curve by $\frac{1}{2}$ as $\omega \rightarrow \infty$ on the Onsager model.

It is just possible this might be found experimentally, but small deviations of this order could be attributed to many factors.

It can definitely be stated that the experimental value of ϵ_0 for water is higher than the currently accepted value of 2.0. It is interesting to note that if we substitute $\epsilon_0 = 5.5$ in Onsager's equation instead of $\epsilon_0 = 2.0$, the calculated value of the dipole moment is found to be 1.5 Debye units.

The accepted value of μ_0 is 1.87 in the vapour phase, and it is therefore apparent that if the high value for ϵ_0 is accepted, Onsager's formula gives, unexpectedly, a value of $\epsilon_s = 130$ which is greater than that found experimentally.

However, this large value for ϵ_0 is not without its own difficulties; the value of ϵ_0 for ice is 3, so that one might expect a considerable decrease in dielectric constant when water freezes. Also, if one substitutes the high value of ϵ_0 in equations (3) and (9), one finds a ratio $\mu/\mu_0 = 2.6$. Since μ_0 is already 1.87 Debye units this would mean so great an electronic polarization that γ could no longer be regarded as constant.

The only possibility is to leave μ throughout the equations, which would finally give

$$\epsilon - n^2 = \frac{4\pi}{V} \cdot \frac{3(2\epsilon_s + n^2)}{2\epsilon_s + 1} \cdot \frac{\mu^2}{3kT(1 + j\omega\tau)} \cdot \frac{\epsilon}{2\epsilon + 1}.$$

It is possible that the anomalous dielectric properties of "associated" liquids are due to the fact that the atomic and electronic polarizations are very much higher than is generally assumed. It can be seen qualitatively that association might produce high polarizabilities.

It is of interest to test the Debye relation $\tau = 8\pi\eta a^3/2kT$ experimentally. We find for water and heavy water a fair approximation to the proportionality of τ and η/T , which is shown in figure 8. It appears that a power law slightly different

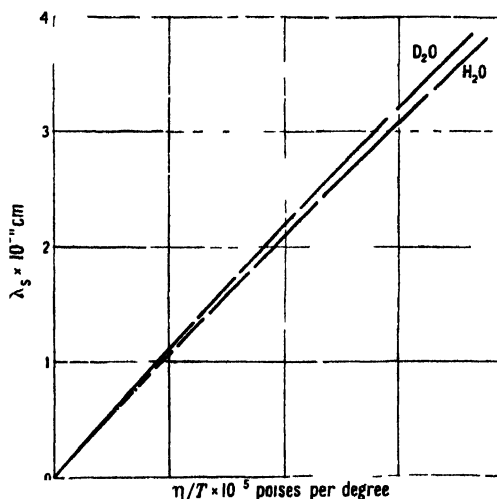


Figure 8. Relation between relaxation time and viscosity

from unity is obeyed; logarithmic plotting shows it to be a 0.97 law, from which extrapolation will give the real relaxation time at any temperature for which η is shown (e.g. for super-cooled water). Calculation of τ on the accepted basis of $a = 1.38 \text{ \AA}$. gives a value of $\tau = 0.85 \times 10^{-11}$, in good agreement with the experimental value of 0.9×10^{-11} .

Further, the ratio of viscosities and of relaxation wavelengths for water and heavy water show fair agreement over a temperature range, as can be seen from table 6 (viscosities from Lewis 1933).

Table 6

Smoothed values of:

T (°C.)	$\lambda_{D_2O} / \lambda_{H_2O}$ ± 2%	$\eta_{D_2O} / \eta_{H_2O}$
10	1.30	1.29
20	1.27	1.25
30	1.24	1.22
40	1.21	1.19*

* Extrapolated

The close numerical agreement between the observed and calculated values of τ is surprising, and probably fortuitous, but the close dependence of τ upon viscosity, both in water and heavy water, seems to us significant. It is evident that viscous flow and dipole orientation are influenced by the same factors, and can thus, in

some ways, be said to favour a hole theory such as that of Eyring (Glasstone, Laidler and Eyring 1941). If one accepts Eyring's result

$$\frac{1}{\tau} = \frac{kT}{h} \cdot e^{\Delta S/R} \cdot e^{-\Delta H/RT},$$

connecting the thermodynamic properties of the activation process involved in re-orientation, one can calculate ΔH from the observed temperature coefficient of τ . The value obtained is about 5 kcal gm. mol. at 0 °C., which is roughly the same as the surface energy. The observed value of ΔH also falls with rising temperature as the surface energy is known to do.

ACKNOWLEDGMENTS

The work described in this paper has been carried out on behalf of the Director of Physical Research, Admiralty, and the authors wish to record their thanks for permission to publish. The authors are grateful to Lord Cherwell for kindly extending to them the facilities of the Clarendon Laboratory. They also wish to thank D.S.I.R. from whom one of them (D.M.R.) has received a grant, and Mr. W. Stonard whose skill in the construction of the absorption cells and other equipment has been of great value.

REFERENCES

- ABADIE, P., and GIRARD, J., 1946, *Trans. Faraday Soc.*, Discussion on Dielectrics.
 BERNAL, J. D., and FOWLER, R. H., 1933, *J. Chem. Phys.*, **1**, 515.
 BOTTYCHER, C. J. F., 1942, *Physica*, **9**, 927.
 CARTWRIGHT, C. HAWLEY, 1936, *J. Chem. Phys.*, **4**, 413.
 CLAYTON, R. J., HOUDIN, J. L., LAMONT, H. R. L. and WILLISIAW, W. E., 1946, *J. Instn. Elect. Engrs.*, Part III, **93**, 97.
 COLLIE, C. H., HASTED, J. B., and RITSON, D. M., 1948, *Proc. Phys. Soc.*, **60**, 71.
 COLLIE, C. H., RITSON, D. M., and HASTED, J. B., 1946, *Trans. Faraday Soc.*, **42A**, 129.
 CONNOR, C. P., and SMYTH, W. P., 1943, *J. Amer. Chem. Soc.*, **65**, 382.
 DEBYE, P., 1929, *Polar Molecules* (Chemical Catalogue Co.), p. 92.
 DIVILKOWSKI and MASCH, 1940, *C.R. Acad. Sci., U.S.S.R.*, **27**, 801.
 DORSEY, N. E., 1940, *Properties of the common water-substance* (Reinhold Publishing Co.).
 ESAU, A., and BÄZ, G., 1937, *Phys. Z.*, **38**, 74.
 FOWLER and GUGGENHEIM, 1939, *Statistical Thermodynamics* (Cambridge).
 GLASSTONE, S., LAIDLER, K. J., and EYRING, H., 1941, *Theory of Rate Processes* (McGraw-Hill Co.).
 HACKEL, W., and WIEN, M., 1937, *Phys. Z.*, **38**, 767.
 JACKSON, WILLIS, 1946, *Trans. Faraday Soc.*, Discussion on Dielectrics.
 KEBBEL, W., 1939, *Hochfrequenztech. u. Elektroakust.*, **53**, 81.
 KNERR, H. W., 1937, *Phys. Rev.*, **52**, 1054.
 LEWIS, G. N., 1933, *J. Amer. Chem. Soc.*, **55**, 4730.
 PENROSE, R. P., 1946, *Trans. Faraday Soc.*, Discussion on Dielectrics.
 RUBENS, H., 1915, *Verh. dtsh. phys. Ges.*, **17**, 315.
 SAXTON, J. A., and LANE, J. A., 1947, *Phys. Soc., Meteorological Factors in Radio-Wave Propagation*, p. 278.
 SLEVOGT, 1939, *Ann. Phys., Lpz.*, **36**, 147.

Quantum Mechanical Calculation of the Heat of Solution and Residual Resistance of Gold in Silver

By KUN HUANG

H. H. Wills Physical Laboratory, University of Bristol

MS. received 7 July 1947

ABSTRACT. An attempt is made to calculate the heat of solution of gold in silver on the basis of the quantum theory of metals. These two metals are chosen because they have the same atomic volume, and therefore are the simplest case.

The steps in the argument are as follows: Suppose that a gold atom replaces a silver atom in the lattice. Then, to a certain approximation, one can represent the substitution of the silver ion by a gold ion, in its effect on the electrons, by a "potential hole" of depth ΔE and radius r_0 . This potential hole will alter the energy of the conduction electrons. To a first approximation the change in energy is just ΔE , which would give zero heat of mixing. A second-order term of order $(\Delta E)^2/E_F$ always gives a positive heat of mixing; E_F is here the Fermi energy. This term is calculated exactly by wave-mechanical methods; it gives 0.69 ev. per atom. The same calculation shows, however, that there is a concentration of charge in the gold atom in excess of that in the surrounding silver atoms; this alters the potential in which the electrons move, so that a self-consistent calculation is required to obtain the true energy. For this the labour required would be almost prohibitive; therefore we use instead the Thomas-Fermi method and obtain 0.45 ev. We thus find 0.15 ev. per atom for the heat of solution, which compares well with the observed value 0.13 ev.

With the help of the potential obtained with the Thomas-Fermi method the residual resistance of gold in silver is found to be 0.16 microhm cm. for 1% solution. The considerable discrepancy as compared with the experimental value 0.38 seems closely connected with very similar discrepancies found in other theoretical work on temperature resistance of the noble metals.

§ 1. INTRODUCTION

IN many applications of the electron theory of metals to alloys, it has been found possible to treat the different constituent atoms merely as sources supplying different numbers of metallic electrons which belong to the metal as a whole. The irregularities created by the foreign atoms have been entirely ignored and the electronic band structure has been treated in the same way as in pure metals. An exception is the work on residual resistance of alloys, for the residual resistance is a direct result of the local irregularities in the potential field caused by the difference in the nature of the constituent atoms. However, most of this work is essentially formal and succeeds only in correlating certain facts, and leaving the question of the nature of the irregularities untouched. The only attempts at quantitative estimation of absolute magnitudes of residual resistance seem to be those made by Mott (1936). An attempt to obtain some idea of cohesion of alloys with the simple model employed by Mott in his work on resistance has naturally led us to a closer examination of the irregularities in the potential surrounding the foreign atoms. It is the purpose of the present paper to investigate their nature more deeply and give a semi-quantitative discussion of the cohesive energy and resistance. For our actual discussion, the alloy Ag-Au will be used, because the constituent atoms have practically the same radii, so that the effects due to distortion

of the lattice will not be present. Similar effects to those discussed here will occur also in Cu-Ag, Cu-Au, but will be complicated by effects of distortion both in the consideration of resistance and energy.

The sum of the sublimation energy and the ionization energy of the atoms in a metal will be called the *binding energy* of the metal; this is the energy necessary to separate the individual ionic cores and valence electrons into free ions and electrons. Then the binding energy of a monovalent pure metal with "almost-free electrons" can be put rather crudely as

$$W = -\sum_i \left(E^0 + \frac{\hbar^2 \mathbf{k}_i^2}{2m} \right), \quad \dots\dots(1)$$

E^0 is a constant for the metal concerned, and \mathbf{k}_i is the wave number vector of the i th electron, so that the second term alone is just the Fermi energy of the electrons. Formula (1) can be derived theoretically (e.g. Seitz 1940, p. 345) when, owing to correlation, only one electron is in each atomic sphere. $E^0 + \hbar^2 \mathbf{k}_i^2 / 2m$ is thus the eigenvalue of the i th state in the Hartree sense. (By adding these eigenvalues the mutual electronic repulsive energy is counted twice, and this serves very conveniently to take care of the repulsion between the ions). Thus E^0 appears as the constant potential energy of the Sommerfeld model. Mott (1936) has actually shown that to the same approximation this simple interpretation can be extended to alloys. Thus the difference in the ionic field, say, of a gold atom in a silver lattice can be regarded as a change in the potential energy from E_{Ag}^0 to E_{Au}^0 in the atomic sphere surrounding the gold ion. This idea is of importance in making possible at least a semi-quantitative treatment of effects involving such irregularities in the potential field. However, simple consideration will show that it is necessary to elaborate the square-hole potential, before it can be used as the effective field in which the electrons move.

Consider, for instance, an Au atom in an Ag lattice. As the atomic radii are the same, the Fermi energies are the same. From (1) it follows that

$$E_{\text{Ag}}^0 - E_{\text{Au}}^0 = W_{\text{Au}} - W_{\text{Ag}} - 2.60 \text{ ev.}, \quad \dots\dots(2)$$

where the values of the binding energy are obtained from the sublimation energy and the atomic ionization energy (Landolt Bornstein 1936). If the square-hole potential

$$U(r) = E_{\text{Ag}}^0 \quad r > r_0 (= 1.59 \times 10^{-8} \text{ cm.}) \\ = E_{\text{Ag}}^0 - 2.60 \quad r < r_0$$

is taken as the effective field in which the individual electrons move, where the Au atom is supposed to be situated at the origin and r_0 is the atomic radius, the eventual distribution of electrons can be obtained very simply by the Thomas-Fermi method with the particles treated as non-interacting. The number of electrons in any volume element $d\tau$ is given by

$$\rho d\tau = (8\pi/3\hbar^3) [2m(E_m - U(r))]^{3/2} d\tau, \quad \dots\dots(3)$$

where E_m is the energy of the highest level filled. Let ρ_0 be the density of electrons in pure Ag. Then if the metal is taken large enough, obviously

$$\rho_0 = (8\pi/3\hbar^3) [2m(E_m - E_{\text{Ag}}^0)]^{3/2},$$

that is, the constant density outside the potential hole is identical with that of the

pure Ag metal and $E_m - E_{\text{Au}}^0$ is equal to its maximum Fermi energy: 5.5 ev. From (3) it follows that the ratio of the density for $r < r_0$ in excess of ρ_0 to ρ_0 is given by

$$\frac{(5.5 + 2.6)^{3/2} - (5.5)^{3/2}}{(5.5)^{3/2}} = 0.78,$$

showing a considerable accumulation of charge at the Au atom. The electrostatic potential produced by the excess of charge can be calculated directly. A curve (C) giving the potential energy of an electron caused by this potential together with a curve (A) showing $(\rho - \rho_0)/\rho_0$ is plotted in figure 1. The energy is seen to rise above 10.0 ev. at the centre of the Au atom. The use of the simple square-hole potential alone as the field in which the electrons move, therefore, does not lead to a self-consistent result

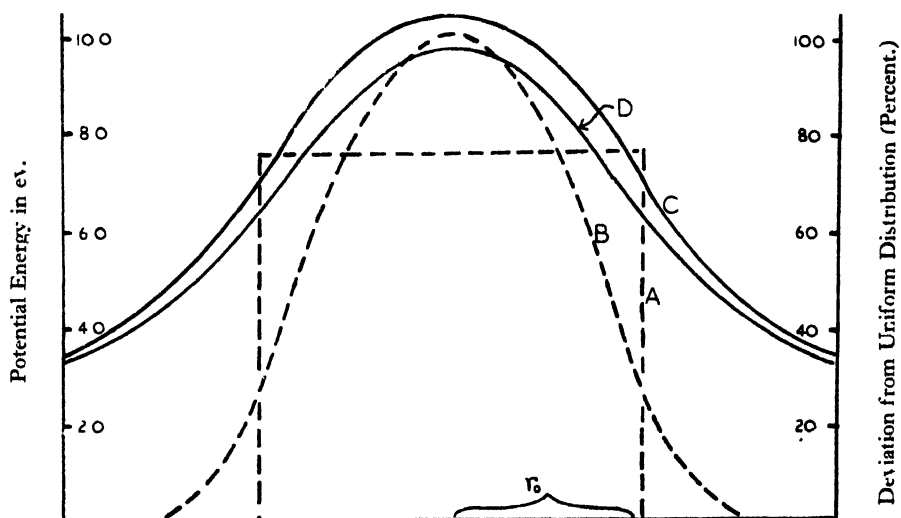


Figure 1.

- A = $\Delta\rho/\rho_0$ by Thomas-Fermi method.
- B = $\Delta\rho/\rho_0$ by wave mechanics.
- C = Potential energy due to $\Delta\rho$ by Thomas-Fermi method.
- D = Potential energy due to $\Delta\rho$ by wave mechanics.

Qualitatively it is clear that when the electronic interaction is properly considered, the potential-hole will be negatively charged due to the preferential accumulation of electrons, and its immediate neighbourhood will be positively charged to screen off the field produced by the negative centre. The rather intricate nature of this charge arrangement suggests the method of Thomas-Fermi (with interacting particles) as the suitable method to treat the problem, at least as a first approximation. To avoid excessive complication the exchange and correlational effects will not be considered, so that the Thomas-Fermi method can be used in its original form. This procedure amounts to ignoring the change in the exchange and correlational effects brought about by the change in electron density.

Since it is difficult to predict the accuracy of the Thomas-Fermi method on *a priori* grounds, the next section will be devoted to finding exact wave functions for electrons in the potential hole, and the results obtained will be compared.

Further developments in §3 concerning energy changes of these cases will be important for later discussions of the energy of cohesion in §5.

§2. EXACT WAVE FUNCTIONS AND CHARGE DISTRIBUTION FOR NON-INTERACTING ELECTRONS

For the present purpose it is convenient to suppose the metal to be in the form of a large sphere of silver of radius R with a single gold atom situated at its centre. The normalized wave functions of the electrons have the general form

$$N_{kl} \frac{G_{kl}(r)}{r} \Phi_{lm}(\theta\phi), \quad \dots\dots (4)$$

where $\Phi_{lm}(\theta\phi)$ is the normalized spherical harmonic for a state with orbital quantum number l and azimuthal quantum number m . N_{kl} is the normalization factor for G_{kl} which satisfies the radial wave equation (e.g. Mott and Massey 1933, p. 22)

$$\begin{aligned} \frac{d^2 G_{kl}}{dr^2} + \left(k'^2 - \frac{l(l+1)}{r^2} \right) G_{kl} &= 0 \quad r < r_0, \\ \frac{d^2 G_{kl}}{dr^2} + \left(k^2 - \frac{l(l+1)}{r^2} \right) G_{kl} &= 0 \quad r > r_0, \end{aligned} \quad \dots\dots (5)$$

where
$$\frac{\hbar k'^2}{2m} = E - E_1, \quad \frac{\hbar^2 k^2}{2m} = E - E_2$$

denote respectively the kinetic energies in the two regions. Here, for simplicity, E_1, E_2 will be written in place of E_{Au}^0 and E_{Ag}^0 .

The solution of (5) that is regular at the origin is

$$\begin{aligned} G_{kl}(r) &= A \sqrt{r} J_{l+1/2}(k'r) \\ &\quad - B \sqrt{r} J_{l+1/2}(kr) + C \sqrt{r} J_{l-1/2}(kr) \quad r > r_0, \end{aligned} \quad \dots\dots (6)$$

$J_{l+1/2}, J_{l-1/2}$ are the Bessel functions of half odd-integer orders. Since G_{kl} and dG_{kl}/dr are continuous at $r=r_0$, the ratio of the constants A, B, C , is fixed so that only a multiplying constant is left arbitrary. To define this, let $G_{kl}(r)$ have the asymptotic form at infinity

$$G_{kl}(r) \sim \left(\frac{2}{\pi k} \right)^{1/2} \sin(kr - \frac{1}{2}l\pi + \eta_{kl}).$$

Then if the radius of the sphere R is taken sufficiently large,

$$\begin{aligned} (N_{kl})^{-2} &\simeq \frac{2}{\pi k} \int_0^R \sin(kr - \frac{1}{2}l\pi + \eta_{kl}) dr \\ &\simeq \frac{R}{\pi k}, \end{aligned} \quad \dots\dots (8)$$

which determines N_{kl} .

For the state represented by (4) to be a stationary state, the condition

$$G_{kl}(R) = 0$$

has to be fulfilled. When R is large (7) may be used. Thus

$$kR = n\pi - \left(\eta_{kl} - \frac{l\pi}{2} \right), \quad \dots\dots (9)$$

where n is an integer. From (9) it follows that the number of states with definite values of l , m , and with k in the interval k to $k + dk$, is given by

$$\frac{2R}{\pi} dk. \quad \dots\dots(10)$$

The factor 2 allows for the two states of spin. With (8) and (10) the electron density can be written down directly in terms of $G_{kl}(r)$. Since the solution corresponding to G_{kl} , where the gold atom is replaced by a silver atom is just $\sqrt{r}J_{l+1/2}(kr)$, which satisfies (7), it is seen that the change in density is given by

$$\begin{aligned} \Delta\rho &= 2 \sum_l \left\{ \sum_{m=-l}^l \Phi_{lm}^*(\theta\phi) \Phi_{lm}(\theta\phi) \right\} \int_0^{k_m} \left(\frac{G_{kl}^2(r)}{r^2} - \frac{J_{l+1/2}^2(kr)}{r} \right) k dk \\ &= \frac{1}{2\pi} \sum_l (2l+1) \int_0^{k_m} \left(\frac{G_{kl}^2(r)}{r^2} - \frac{J_{l+1/2}^2(kr)}{r} \right) k dk, \quad \dots\dots(11) \end{aligned}$$

k_m is the wave number corresponding to the highest filled state. Owing to the complicated natures of the coefficients A , B , C , involved in G_{kl} , the integral can only be evaluated numerically. The result is shown in figure 1, curve B. The Coulomb energy of an electron due to this excess charge density has been worked out from this curve by numerical methods and the result is represented by curve D in the same figure. Owing to the discontinuous nature of the density function as given by the Thomas-Fermi method, the general agreement between A and B is as good as can be expected. The potential is not discontinuous, and the agreement between C and D is remarkably close. Since the case treated actually involves a discontinuous drop in the potential, and since nevertheless the simple Thomas-Fermi method still yields reasonably good results, it seems reasonable to believe that the method will in general lead to fairly reliable conclusions.

§ 3 ENERGY CHANGES IN THE CASE OF HYPOTHETICAL NON-INTERACTING ELECTRONS

It is well known that the Thomas-Fermi method fails generally to take account of the kinetic energy associated with rapid variation in electron density and gives too low a value for the energy when rapid changes in potential are involved. As the square hole potential will still be used in the discussion on cohesive energy, it is important to be able to estimate the inaccuracy caused by its discontinuous change in potential, and to correct for it. The most direct, and probably the best, course is to calculate the energy change caused by the presence of the Au atom in the hypothetical case that has been treated by both methods and use the difference to correct for the inaccuracy in later discussions.

In the Thomas-Fermi method the average energy of an electron in any particular volume element is taken to be the sum of the potential energy at this point and $3/5$ the maximum kinetic energy. As ρ_0 corresponds to just one electron per atom, it follows from § 1 that the number of electrons in the gold sphere is given by $(8.1/5.5)^{3/2}$, and the total energy of these electrons is

$$\left(\frac{8.1}{5.5} \right)^{3/2} \left(\frac{3}{5} \times 8.1 + E_1 \right). \quad \dots\dots(12)$$

Of these, for a pure metal, one electron is normally in the same sphere and the others are to be regarded as drawn from the top of the Fermi distribution of the

whole metal, so that the other electrons might be regarded as not perturbed by the presence of the Au atom. Therefore the energy of the "same" electrons, before the Au atom ion replaces the Ag ion, is

$$\frac{3}{5} \times 5.5 + \left[\left(\frac{8.1}{5.5} \right)^{3.2} - 1 \right] \times 5.5 + \left(\frac{8.1}{5.5} \right)^{3.2} \times E_2. \quad \dots (13)$$

Subtracting (13) from (12), and using the value $E_2 - E_1 = 2.60$ ev. as given by (2), the following value for the energy change is obtained

$$-3.59 \text{ ev.}$$

A more significant quantity is given by

$$-3.59 - (E_1 - E_2) = -0.99 \text{ ev.}, \quad \dots (14)$$

which is the energy change accompanying the transfer of one gold atom from bulk gold to pure silver in this hypothetical case.

The energy change corresponding to the rigorous treatment in §2 can be expressed in terms of the phase shifts η_{kl} introduced in (7). We recall that the solution corresponding to $G_{kl}(r)$, before the Au atom replaces the Ag at the origin is $\sqrt{r} J_{l+1/2}(kr)$, so that the phase shifts are originally zero. It then follows from (9) that if R is taken to be very large, the value of k corresponding to the same quantum number n is altered by the infinitesimal amount

$$dk = -\eta_{kl}/R,$$

when the Au atom is introduced. Accordingly, the change in energy is given by

$$d\left(\frac{\hbar^2 k^2}{2m}\right) = -\frac{\hbar^2 k}{m} \frac{\eta_{kl}}{R},$$

which is independent of the azimuthal quantum number, so that on summing over all the filled states with the help of the density of states given by (10), the total energy change is obtained in the form of a sum of integrals

$$-\left(\frac{\hbar^2}{m\pi}\right) \Sigma 2(2l+1) \int_0^{k_m} k \eta_{kl} dk. \quad \dots (15)$$

In connection with the evaluation of the expression (15), a simple and rather interesting identity may be deduced, which makes it possible to obtain a result of the desired degree of accuracy without numerical work involving terms with $l \geq 2$, so that the necessary values of η_{kl} can be taken directly from results obtained in working out the last section.

It is well known from the theory of collisions that an approximation to η_{kl} is given by (Mott and Massey 1933, p. 28)

$$\eta_{kl} \cong -\frac{\pi m}{\hbar^2} \int_0^\infty U(r) (J_{l+1/2}(kr))^2 r dr, \quad \dots (16)$$

where $U(r)$ is the perturbing potential which vanishes at infinity, so that

$$U(r) = -\Delta E = -(E_2 - E_1) \quad r < r_0 \\ = 0 \quad r > r_0 \quad \dots (17)$$

in the present case. Substituting (16), (17) into (15), the following approximate value for the change of energy is obtained:

$$-\Delta E \Sigma 2(2l+1) \int_0^{k_m} k dk \int_0^{r_0} (J_{l+1/2}(kr))^2 r dr. \quad \dots (18)$$

On the other hand, on using the expansion (Watson 1944)

$$e^{ikr} = \left(\frac{\pi}{2kr}\right)^{1/2} \sum_{l=0}^{\infty} i^l (2l+1) P_l(\cos \theta) J_{l+1/2}(kr),$$

it is easily established after integration over θ that

$$\left. \begin{aligned} \frac{4\pi r_0^3}{3} &= \int_{\text{atomic sphere}} e^{i\mathbf{k} \cdot \mathbf{r}} \times e^{ikr} dr \\ &= \frac{\pi^2}{k} \sum 2(2l+1) \int_0^{r_0} r (J_{l+1/2}(kr))^2 dr. \end{aligned} \right\} \dots\dots (19)$$

On substituting (19) in (18), the following identity is obtained:

$$\left. \begin{aligned} -\Delta E &\sum 2(2l+1) \int_0^{k_m} dk \int_0^{r_0} (J_{l+1/2}(kr))^2 r dr \\ &= -\Delta E \left(\frac{4}{9\pi}\right) (k_m r_0)^3 = -\Delta E, \end{aligned} \right\} \dots\dots (20)$$

where use is made of the fact that the distribution corresponds to a density of one electron per atomic sphere. (20) states the fact that when the approximation (16) for η_{kl} is used, the energy change obtained is just $-\Delta E$, the value that would have been obtained if the first-order perturbation theory had been used.

We use this result in the following way. We desire to find the change in energy using exact values of η_{k0} , η_{k1} only. With the help of (20), the change of energy can then be written as

$$\begin{aligned} -\Delta E + \Delta E \left[2 \int_0^{k_m} k dk \int_0^{r_0} (J_{1/2}(kr))^2 r dr + 6 \int_0^{k_m} k dk \int_0^{r_0} (J_{3/2}(kr))^2 r dr \right] \\ - \frac{h^2}{2m} \left[2 \int_0^{k_m} k \eta_{k0} dk + 6 \int_0^{k_m} k \eta_{k1} dk \right], \end{aligned} \dots\dots (21)$$

which corresponds to (15), but with values of η_{kl} given by (16) for $l \geq 0$. The integrals in the second term can be expressed in terms of the Si functions (Jahnke and Emde 1945) and the integrals in the third term are evaluated by numerical integrations. The result is

$$-\Delta E + \Delta E(0.957) - 3.18 = -3.29 \text{ ev.}$$

The first two terms together give the total contribution of all states with $l \geq 2$. It has the very small value 0.11 ev.; the assumption that deviations from (16) could be ignored for states with $l \geq 2$ is therefore justified. The more significant quantity corresponding to (14) of the Thomas-Fermi method (the heat of mixing) is in this case

$$-3.29 + \Delta E = -0.69 \text{ ev.} \dots\dots (22)$$

Comparison of (22) with (14) shows that the energy value given by the Thomas-Fermi method is too low by 0.30 ev. Although the discrepancy is not very great, considering the abrupt change in the potential, it is, however clear that the conclusions regarding cohesive energy developed in later sections cannot be expected to be more than semi-quantitative. For the ultimate quantity of interest (heat of mixing) corresponding in the real case to (22) and (14) in the above hypothetical case should be of the order 0.1–0.2 ev. So the correction of 0.30 ev. is by comparison very considerable.

§ 4. APPLICATION OF THE THOMAS-FERMI METHOD TO THE ACTUAL CASE

The considerations of § 1 show that it is necessary to take electronic interaction properly into account to obtain a self-consistent result. The square-hole potential describes the difference of the ionic field within the atomic sphere of the gold atom from that of the rest; so that if the electron cloud were uniform, it would be the effective potential in which the electrons move. Therefore, in order to treat the actual case, a natural procedure would be to take the potential as the superposition of the square-hole potential and that produced by a uniform space distribution of positive charge that would neutralize the electronic charge completely if the latter were uniformly distributed. Let $U(r)$ be the square hole potential as already given by (17) and $V(r)$, $-e$ be respectively the electrostatic potential and the electronic charge. The density of electrons at any point is then $(8\pi/3h^3)[2m(E_m - U + V)]^{3/2}$, which leads to the following Poisson's equation for V :

$$\nabla^2 V = -4\pi\rho = \frac{32\pi^2 e}{3h^3} [2m(E_m - U + eV)]^{3/2} \cdot \frac{3e}{r_0^3}, \quad \dots\dots (23)$$

for the positive charge density is $e(4\pi r_0^3/3)^{-1}$. Equation (23) can be simplified by introducing the dimensionless quantities

$$\left. \begin{aligned} g &= (E_m - U(r) + eV(r)) / E_F, \\ \xi &= r / (r_0 E_F / 3e^2)^{1/2} r_0, \end{aligned} \right\} \dots\dots (24)$$

where

$$E_F = \left(\frac{9\pi}{4} \right)^{2/3} \frac{h^2}{2mr_0^2}$$

is the maximum kinetic energy of a free electron gas of the same average density. In terms of (24), (23) becomes

$$\frac{1}{\xi^2} \frac{d}{d\xi} \left(\xi^2 \frac{dg}{d\xi} \right) = g^{3/2} - 1. \quad \dots\dots (25)$$

Introducing f defined by

$$f(\xi) = \xi g(\xi), \quad \dots\dots (26)$$

(25) reduces to

$$\frac{d^2 f}{d\xi^2} = \xi \left[\left(\frac{f}{\xi} \right)^{3/2} - 1 \right]. \quad \dots\dots (27)$$

Since the change in electron density is small compared with the average density, it is advantageous to introduce $\Delta(\xi)$ defined by

$$f(\xi) = \xi + \Delta(\xi). \quad \dots\dots (28)$$

In all cases treated, Δ/ξ can be treated as small and it is sufficient to solve (27) to the second approximation. We thus substitute (28) in (27) and expand $f^{3/2}$ in powers of Δ/ξ to the third term. The result is

$$\frac{d^2 \Delta}{d\xi^2} = \frac{3}{2} \Delta + \frac{3}{8} \frac{\Delta^2}{\xi}. \quad \dots\dots (29)$$

Before solving (29) it is useful first to examine the boundary conditions to be satisfied by Δ . As Δ must tend to zero as ξ tends to infinity, it is clear from (29) that Δ must have the asymptotic form

$$\Delta \sim A e^{-\sqrt{3/2} \xi}, \quad \dots\dots (30)$$

where A is an arbitrary constant.

At the origin, as the charge density must be finite, it follows directly from Gauss's law that

$$\frac{dV}{dr} \rightarrow 0 \quad \text{as} \quad r \rightarrow 0,$$

and hence that $\frac{dg}{dr} \rightarrow 0$.

Therefore it follows that

$$\Delta \rightarrow \xi \frac{d\Delta}{d\xi} \quad \text{and} \quad \frac{d\Delta}{d\xi} \rightarrow g-1.$$

Since $g-1$ must remain finite, it is necessary that

$$\Delta(\xi) \rightarrow 0 \quad \text{as} \quad \xi \rightarrow 0. \quad \dots\dots(31)$$

Therefore in the solution at the origin only one arbitrary constant remains, defined by

$$B = \left(\frac{d\Delta}{d\xi} \right)_{\xi=0},$$

which, together with the arbitrary constant A in the asymptotic solution at infinity, is to be determined by the boundary conditions on the surface of the atomic sphere $\xi = \xi_0$. On this surface the electrostatic potential and field must be continuous. It follows that

$$\begin{aligned} \Delta_2(\xi_0) - \Delta_1(\xi_0) &= \xi_0(g_2(\xi_0) - g_1(\xi_0)) = \xi_0 \frac{(U_1(\xi_0) - U_2(\xi_0))}{E_F} = \xi_0 \frac{E_1 - E_2}{E_F}, \\ \left(\frac{d\Delta_2}{d\xi} \right)_{\xi_0} - \left(\frac{d\Delta_1}{d\xi} \right)_{\xi_0} &= (g_2(\xi_0) - g_1(\xi_0)) + \xi_0 \left(\left(\frac{dg_2}{d\xi} \right)_{\xi_0} - \left(\frac{dg_1}{d\xi} \right)_{\xi_0} \right) \\ &= U_1(\xi_0) - U_2(\xi_0) = \frac{E_1 - E_2}{E_F}, \quad \dots\dots(32) \end{aligned}$$

where the suffixes 1 and 2 indicate the limiting value for $\xi < \xi_0$ and $\xi > \xi_0$ respectively. (32) will determine A and B , which are left arbitrary in (30) and (31).

Let $\phi(\xi)$ be a solution of the homogeneous part of equation (29), so that

$$\frac{d^2\phi}{d\xi^2} = \frac{3}{2}\phi. \quad \dots\dots(33)$$

Construct the integral

$$\int_{\xi_1}^{\xi_2} \left(\phi \frac{d^2\Delta}{d\xi^2} - \Delta \frac{d^2\phi}{d\xi^2} \right) d\xi.$$

A simple application of Green's theorem leads, with the help of (29) and (3), to the following relation:

$$\left(\phi \frac{d\Delta}{d\xi} - \Delta \frac{d\phi}{d\xi} \right) \Big|_{\xi_1}^{\xi_2} = \frac{3}{2} \int_{\xi_1}^{\xi_2} \phi \frac{\Delta^2}{\xi} d\xi, \quad \dots\dots(34)$$

which makes it possible to give explicitly the value of Δ and $d\Delta/d\xi$ at any value of ξ , both for $\xi > \xi_0$ and $\xi < \xi_0$, in terms of the constants A , B in (30) and (31) and the integral involving the small quantity of second order Δ^2/ξ^2 . For the region $\xi > \xi_0$, it is only necessary to take $\xi_2 = \infty$, $\xi_1 = \xi$ and put ϕ equal to $e^{\sqrt{3/2}\xi}$ and $e^{-\sqrt{3/2}\xi}$ successively. The limit on the left-hand side at ξ_2 is determined with the help of (30) and two linear equations are thus obtained with Δ and $d\Delta/d\xi$ as unknown

and may be readily solved. For $\xi < \xi_0$ the procedure is similar, the only difference being that ξ_1 is taken to be at the origin and (31) is used accordingly. A first approximation to Δ involving only the constant A or B may be obtained by leaving out the integral involving Δ^2/ξ . When these first approximations are used for Δ in the integral, Δ and $d\Delta/d\xi$ are obtained in an explicit form to the second approximation. The following are these solutions:

$$\xi < \xi_0 \quad \left. \begin{aligned} \Delta &= \frac{1}{2\alpha} \left\{ 2\alpha A e^{-\alpha\xi} + \frac{3A^2}{8} [e^{\alpha\xi} Ei(-3\alpha\xi) - e^{-\alpha\xi} Ei(-\alpha\xi)] \right\} \\ \frac{d\Delta}{d\xi} &= \frac{1}{2} \left\{ -2\alpha A e^{-\alpha\xi} + \frac{3A^2}{8} [e^{\alpha\xi} Ei(-3\alpha\xi) + e^{-\alpha\xi} Ei(-\alpha\xi)] \right\} \end{aligned} \right\} \dots\dots (35)$$

$$\xi > \xi_0 \quad \left. \begin{aligned} \Delta &= \frac{1}{2\gamma} \left\{ 2B \sinh \gamma\xi + \frac{B^2}{4} \left[e^{\gamma\xi} \int_0^{\alpha\xi} e^{-x} \frac{\sinh^2 x}{x} dx - e^{-\alpha\xi} \int_0^{\gamma\xi} e^x \frac{\sinh^2 x}{x} dx \right] \right\} \\ \frac{d\Delta}{d\xi} &= \frac{1}{2} \left\{ 2B \cosh \gamma\xi + \frac{B^2}{4} \left[e^{\alpha\xi} \int_0^{\alpha\xi} e^{-x} \frac{\sinh^2 x}{x} dx - e^{-\alpha\xi} \int_0^{\gamma\xi} e^x \frac{\sinh^2 x}{x} dx \right] \right\} \end{aligned} \right\} \dots\dots (36)$$

where for brevity γ is used for $\sqrt{3/2}$, and

$$Ei(-x) = - \int_x^\infty \frac{e^{-t}}{t} dt.$$

(35) and (36) are not particularly convenient to give the general form of Δ , but they make it possible to determine directly A , B from the boundary conditions (32). Once they are determined, a single numerical integration of equation (27) provides the complete solution in numerical form and serves furthermore to check the accuracy of (35) and (36). The constants determined are

$$A = -5.50, \quad B = 0.1122.$$

The error in Δ , Δ' is less than 1% and is accordingly smaller in f and g . The total effective electronic potential energy as function of ξ calculated from the numerical solution is represented in figure 2. In the same figure is also shown the corresponding distribution of charge. It is seen that the qualitative feature of the charge distribution is just as expected, a negatively charged core surrounded by an equal amount of positive charge in its immediate neighbourhood. The effect on the potential is to fill up the potential hole to a very considerable extent and should tend to reduce the electrical resistance; this will be discussed quantitatively in § 6.

§ 5 HEAT OF SOLUTION OF A DILUTE SOLUTION OF Au IN Ag

The heat of solution is a convenient measure for the cohesion of an alloy. For a dilute solution of Au in Ag, it is the energy change accompanying the transfer of one gold atom from bulk gold into bulk silver. First consider the energy change in the system, when an Ag ion in pure bulk Ag is replaced by an Au ion. The interaction of positive ions is not altered and may be left out of our consideration. If one adds up the energies of the electrons regarded as moving in the permanent field $U = eV$, one counts the Coulomb energy of the electron cloud twice over, so the energy of the system is obtained by subtracting from the sum of the electron energies the Coulomb energy of the electron cloud. Because of the spherical symmetry, for purpose of calculating change of energy it is only necessary

to consider the electrons in a finite sphere of radius R if R is taken large enough. Therefore the energy change is given by

$$\lim_{R \rightarrow \infty} \int \rho \left\{ \frac{3}{5} (E_m - U + eV) + (U - eV) - \frac{e}{2} \left[V - \frac{3e}{r_0^3} \left(\frac{R^2}{2} - \frac{r^2}{6} \right) \right] \right\} dr \\ - \int \rho_0 \left\{ \frac{3}{5} E_F + E_1 - \frac{e}{2} \left[- \frac{3e}{r_0^3} \left(\frac{R^2}{2} - \frac{r^2}{6} \right) \right] \right\} dr. \quad \dots\dots(37)$$

The first integral corresponds to the solution after the Au replaces the Ag ion and the second integral corresponds to the pure Ag. In both integrals, the three terms give respectively the kinetic energy, potential energy and half of the Coulomb energy due to the electron cloud alone. In (37) the Coulomb potential due to the

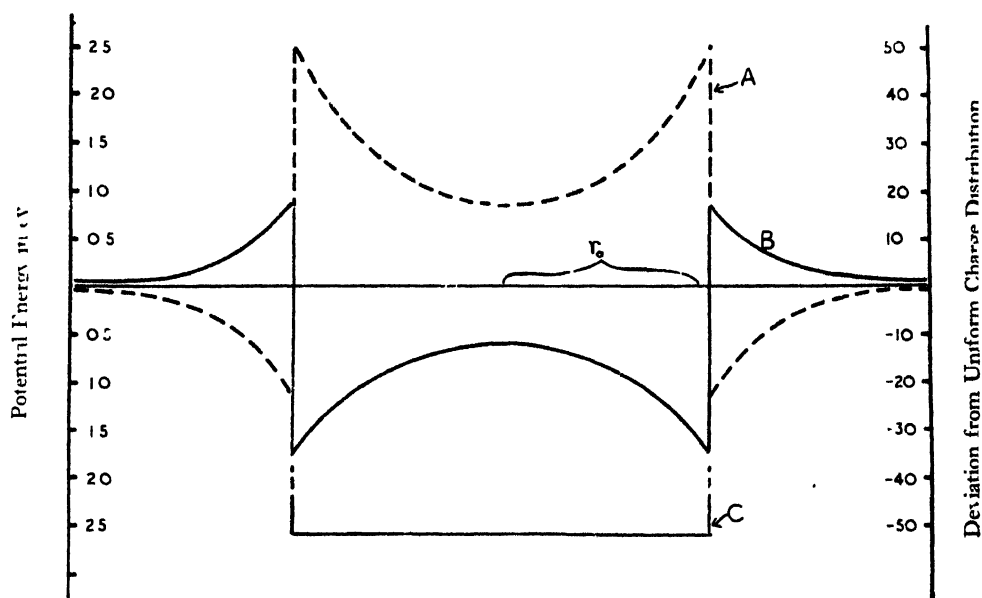


Figure 2

A = Deviation from uniform charge $-\Delta\rho/\rho_0$.

B = Potential energy curve $U - eV$.

C = Original potential hole.

electron cloud alone has been obtained by subtracting from the total Coulomb potential V , the potential produced by the uniform positive charge, namely

$$\frac{3e}{r_0^3} \left(\frac{R^2}{2} - \frac{r^2}{6} \right).$$

Expressing (37) in terms of the dimensionless quantities g and ξ , and making use of the fact that the solution satisfying (30) corresponds to a neutral distribution

$$\lim_{R \rightarrow \infty} \int_0^R (g^{3/2} - 1) dr = 0, \quad \dots\dots(38)$$

(37) becomes

$$3E_F \left(\frac{1}{\xi_0} \right)^3 \lim_{R \rightarrow \infty} \int_0^R \xi^2 d\xi \left\{ g^{3/2} \left(\frac{g}{10} + \frac{U - E_1}{2E_1} + \frac{\xi^2}{12} \right) - \left(\frac{1}{10} + \frac{\xi^2}{12} \right) \right\}, \quad \dots\dots(39)$$

which leads to

$$3E_F \left(\frac{1}{\xi_0} \right)^3 \lim_{R \rightarrow \infty} \int_0^R \left\{ \frac{\xi^2}{10} (g-1) + \frac{\xi^2 (U-E_1)}{2E_F} + \frac{d}{d\xi} \left(\xi \frac{dg}{d\xi} \right) \left(\frac{g}{10} + \frac{U-E_1}{2E_F} + \frac{\xi^2}{12} \right) \right\} d\xi,$$

on account of (25). The integral can be simplified by integration by parts. It is to be noted that both g and U are discontinuous at ξ_0 , so the integration has to be carried out in two sections. Then after a certain amount of manipulation with the help of the boundary conditions satisfied by g , most particularly those at ξ_0 as given by (32), the following expression is obtained:

$$(E_1 - E_2) - 3\xi_0^3 \left\{ \frac{2}{5} (E_2 - E_1) \xi_0^2 \left(\frac{dg}{d\xi} \right)_{\xi_0} + \frac{1}{10} \int_0^{\infty} \xi^2 \left(\frac{dg}{d\xi} \right)^2 d\xi - \frac{3}{5} \int_0^{\infty} \xi^2 (g-1) d\xi \right\}. \quad \dots\dots (40)$$

This expression gives the energy change when a gold atom is substituted for a silver atom.

The most important feature of (40) is the fact that $(E_1 - E_2)$ is naturally separated out from the rest. It is the energy change that would be obtained if the electronic wave functions were not distorted by the potential hole; and at the same time, it is, according to (1), the energy difference between a metallic Au atom and a metallic Ag atom (i.e. average energy per atom in bulk material), so it leaves the rest of (40), namely

$$3\xi_0^3 \left\{ \frac{2}{5} (E_2 - E_1) \xi_0^2 \left(\frac{dg}{d\xi} \right)_{\xi_0} + \frac{1}{10} \int_0^{\infty} \xi^2 \left(\frac{dg}{d\xi} \right)^2 d\xi - \frac{3}{5} \int_0^{\infty} \xi^2 (g-1) d\xi \right\}, \quad \dots\dots (41)$$

as the heat of formation per atom of Au in Ag. This is a very small quantity compared with $(E_1 - E_2)$, so the natural separation of $(E_1 - E_2)$ is of considerable significance in the actual evaluation of the heat of formation. The last integral in (41) contains an integrand which takes on both negative and positive values and nearly cancels out when integrated over the whole range. It is, however, possible to make use of (38) and obtain an approximate but more convenient expression for it of sufficient accuracy. If we write (38) as

$$\begin{aligned} 0 &= \int_0^{\infty} \xi^2 d\xi (g^{3/2} - 1) \\ &= \int_0^{\infty} \xi^2 d\xi \left[\left(1 + \frac{\Delta}{\xi} \right)^{3/2} - 1 \right] \simeq \int_0^{\infty} \left[\frac{3}{2} \frac{\Delta}{\xi} + \frac{3}{8} \left(\frac{\Delta}{\xi} \right)^2 \right] \xi^2 d\xi \end{aligned}$$

or

$$\int_0^{\infty} (g-1) \xi^2 d\xi = \int_0^{\infty} \left(\frac{\Delta}{\xi} \right) \xi^2 d\xi \simeq -\frac{1}{4} \int_0^{\infty} \xi^2 \left(\frac{\Delta}{\xi} \right)^2 d\xi = -\frac{1}{4} \int_0^{\infty} \xi^2 (g-1)^2 d\xi, \quad \dots\dots (42)$$

we introduce an error less than 1% in (41). Thus (41) becomes

$$-3\xi_0^3 \left\{ \frac{2}{5} (E_2 - E_1) \xi_0^2 \left(\frac{dg}{d\xi} \right)_{\xi_0} + \frac{1}{10} \int_0^{\infty} \xi^2 \left(\frac{dg}{d\xi} \right)^2 d\xi + \frac{3}{2} \int_0^{\infty} \xi^2 (g-1)^2 d\xi \right\}. \quad \dots\dots (43)$$

The integrand is now always positive and is convenient for evaluation, and furthermore, it shows quite generally that the heat of solution must always be positive. The use of the numerical solution obtained in the last section gives the value 0.45 ev. Applying the correction to take account of the kinetic energy associated

with the rapid change of potential of the potential hole obtained in §3, the value 0.15 ev. is obtained. The value obtained by extrapolation to infinite dilution of Wachter's data (1932) from measurements made on electrochemical cells is 0.13 ev. The agreement is close. In view of the considerable correction that has to be made, however, the value 0.15 ev. should probably not be taken as more than an indication of the order of magnitude.

§ 6. RESIDUAL RESISTANCE

It is fairly straightforward to work out the residual resistance due to the dissolved Au atoms from the numerical solution of the field obtained in §4 and represented in figure 2. The residual resistance is determined by the scattering of the electrons on the surface of the Fermi distribution (e.g. Mott and Jones 1936, p. 258). If the wave functions of such electrons are of the form given by (4), the equation satisfied by them is (Mott and Massey 1933, p. 22)

$$\frac{d^2 G_l}{d(kr)^2} + \left(1 - \frac{U(r)}{E_F} - \frac{l(l+1)}{(kr)^2}\right) G_l = 0, \quad \dots\dots(44)$$

where $U(r)$ is the potential energy of an electron moving in the effective field, the zero being chosen so that $U(r) \rightarrow 0$ away from the Au ion. As $U(r)$ is only known in numerical form in the present case, it is very much simpler to treat this collision problem by the method of phase shifts (Mott and Massey 1933). If the phase shifts η_l are defined in terms of the asymptotic expression for G_l by (7), the scattering cross-section into a solid angle $d\omega$ is given by

$$I(\theta) d\omega = |f(\theta)|^2 d\omega, \quad \dots\dots(45)$$

where $f(\theta)$ is given by the general expansion

$$f(\theta) = \frac{1}{2ik_m} \sum_{l=0}^{\infty} (2l+1)(e^{2i\eta_l} - 1) P_l(\cos \theta). \quad \dots\dots(46)$$

The resistivity due to Au ions in a dilute solution in which the ions may be regarded as scattering electrons independently of one another, is given in terms of $I(\theta)$ by the formula (Mott 1936)

$$\rho = \frac{h k_m}{e^2} x A, \quad \dots\dots(47)$$

with

$$A = \int_0^\pi (1 - \cos \theta) I(\theta) 2\pi \sin \theta d\theta,$$

where x is the atomic percent. of the Au atoms. Substituting (45), (46) and (47) and integrating over the angle θ with the help of the well known properties of the Legendre polynomials, it is found that

$$\rho = \frac{2h}{e^2 k_m} \sum_{l=0}^{\infty} [(2l+1) \sin^2 \eta_l - 2l \sin \eta_l \sin \eta_{l-1} \cos(\eta_{l-1} - \eta_l)]. \quad \dots\dots(48)$$

Owing to the somewhat larger size of the scattering centre (figure 2) as compared with the potential hole treated in §2 and §3, contributions from $l=2$ term are also considered. η_1 and η_2 are obtained with the first approximation given by (16) and η_0 has been obtained with better approximate methods so that the error is of the order of 1%. The values obtained are

$$\eta_0 = 0.189, \quad \eta_1 = 0.005, \quad \eta_2 = -0.025. \quad \dots\dots(49)$$

In this case η_1 is smaller than η_2 for U is partly negative and partly positive and the maximum of G_1 lies somewhere in between, whereas G_2 has its maximum entirely in the region where U is negative. Using (49) in (48) and converting the result into practical units, the following resistivity for 1% of Au atoms is obtained:

$$0.16 \text{ micro ohm cm.}$$

The agreement with the experimental value 0.38 micro ohm cm. is poor. It is particularly to be noticed that the close agreement of the potentials of the hypothetical case treated in §1 and §2 calculated with the Thomas-Fermi method and the rigorous method indicate clearly that the application of the statistical method to obtain the field should be almost quantitatively correct. It is, however, believed that this discrepancy does not reflect any particular inadequacy of the model used for the dissolved ion. In the first place the resistance depends on the square of the depth of the potential hole, so it is a very sensitive quantity. Secondly, it only depends on a minute fraction of the bulk of the electrons in the metal which are most likely to be affected by the zone structure. Furthermore, it should be remarked that the only serious attempts at calculating theoretically the temperature dependent part of resistivity of the noble metals by Bardeen (1940) and Peterson and Nordheim (1937) show very similar discrepancies with experimental results (Bardeen's result shows almost exactly the same proportional error as the result just obtained, and Peterson and Nordheim's result is too small by a factor 2). It appears therefore reasonable to suppose that all these discrepancies are due to treating the electrons on the surface of the Fermi distribution as free.

§7 CONCLUSION

It is clear from the preceding considerations that when a foreign atom is dissolved in a metal, an appreciable polarization of the electron cloud will in general be caused in its immediate neighbourhood. Either a negatively charged centre is created, surrounded in its immediate neighbourhood by an equal amount of positive charge, or the reverse takes place. Any theory of cohesion must take this fact into account. Owing to the limitation of the method employed, the reasonable agreement of the result obtained in this paper with the observed value could not be looked upon as more than semi-quantitative. It is to be appreciated that the small value of the heat of solution as compared with the primary perturbation in conjunction with the delicate question of polarization makes the problem difficult to treat with great accuracy. Finally, the polarization of the charge tends to fill up irregularities in the potential that would otherwise be there. This reduces the residual resistances caused by the presence of foreign atoms very considerably. In the case treated in the paper, namely, Au in Ag, this leads to a result rather at variance with experiments but at the same time brings it into line with other theoretical works on temperature resistance, indicating therefore a general inaccuracy in treatment rather than particular inadequacy connected with the model employed in this paper to represent a foreign atom.

ACKNOWLEDGMENT

In conclusion, I should like to express my sincere thanks to Professor N. F. Mott, F.R.S., for many discussions and for his kind interest in the work.

REFERENCES

- BARIDEN, J., 1940, *J. Appl. Phys.*, **11**, 88.
 JAHNKE and EMDE, 1945, *Tables of Functions*, 4th edition (New York : Dover Publications).
Landolt Börnstein Tables, 1936, 5th edition, 3rd supplement.
 MOTT, N. F., 1936, *Proc. Camb. Phil. Soc.*, **32**, 281.
 MOTT, N. F., and JONES, H., 1936, *Theory of Properties of Metals and Alloys* (Oxford).
 MOFF, N. F., and MASSEY, H. S. W., 1933, *Theory of Atomic Collisions* (Oxford).
 PETERSON, E. L., and NORDHEIM, L. W., 1937, *Phys. Rev.*, **51**, 335.
 SEITZ, F., 1940, *Modern Theory of Solids* (McGraw Hill).
 WACHTER, A., 1932, *J. Amer. Chem. Soc.*, **54**, 4609.
 WATSON, G. N., 1944, *Theory of Bessel Functions* (Cambridge : University Press).

Electron Momenta in Atoms

By W. E. DUNCANSON * AND C. A. COULSON †

* University College, London. † Physical Chemistry Laboratory, Oxford,
 now at King's College, London

MS. received 12 June 1947

ABSTRACT. Theoretical expressions are obtained for the momentum distribution of the electrons in atoms of different atomic number. This enables density curves to be drawn for the momentum p . The mean momentum, which is also calculated, shows a steady increase with atomic number. Finally, the new calculations of momentum distribution enable earlier work on the shape of the Compton profile to be extended from sodium to potassium, in monatomic form.

§1 INTRODUCTION

THE spatial distribution of electrons in an atom has been fully studied in recent years, and the use of variational methods and self-consistent-field modifications for solving the complete Schrödinger wave equation for an atom have enabled electron density curves to be drawn, in ample and satisfactory agreement with experiments using x-ray scattering. But on the other hand, the velocity distribution is relatively unknown. Yet it is, as Dirac (1947) has shown, equally fundamental. In addition, as Jauncey and DuMond (DuMond 1933) have explained, it is the velocity distribution which determines the shape of the Compton line in x-ray scattering; and, as Hughes has demonstrated both experimentally and theoretically (e.g. Hughes and Mann, 1938) this velocity distribution also describes the inelastic scattering of electrons. It has seemed desirable, therefore, to make explicit calculations of the velocity distribution in simple atoms ($Z < 20$); the results of such calculations are presented in this present paper.

§2. METHOD OF CALCULATION

Following the methods developed by the writers in a series of papers since 1941 (see Duncanson 1943 for a complete list of references), we calculate the momentum p of an electron rather than its velocity. As we use atomic units throughout, the units of momentum and velocity for an electron have the same numerical value, viz. $c/137$. The distribution of p for a single electron is governed by a momentum wave function $\chi(p)$; there is one $\chi(p)$ for every type of atomic

orbital, and $\chi(\mathbf{p})$ is best found from the normalized space wave function $\psi(\mathbf{r})$ by using a fundamental formula of the Dirac transformation theory:

$$\chi(\mathbf{p}) = (2\pi)^{-3/2} \int \exp(-i\mathbf{p}\mathbf{r})\psi(\mathbf{r}) d\mathbf{r}. \quad \dots\dots(1)$$

In this form $\chi(\mathbf{p})$ is already normalized. Associated with $\chi(\mathbf{p})$ is the mean radial distribution function $I(p)$. This function is defined (Duncanson and Coulson 1945, referred to hereafter as D.C.) in such a way that $I(p) dp$ is the probability that the electron has a momentum whose magnitude, independent of direction, lies between p and $p + dp$. Evidently

$$I(p) = \int \chi(\mathbf{p})\chi^*(\mathbf{p})p^2 d\omega_p, \quad \dots\dots(2)$$

where the integration is taken over all values of the solid angle ω_p . In this way $I(p)$ is normalized so that

$$I(p)dp = 1. \quad \dots\dots(3)$$

The mean momentum \bar{p} for this orbital is given by

$$\bar{p} = \int_0^\infty p I(p) dp. \quad \dots\dots(4)$$

Our present calculations start with a presumed knowledge of $\psi(\mathbf{r})$ for each orbit in the atom: (1)–(3) then allow us to determine $\chi(\mathbf{p})$, $I(p)$ and \bar{p} for the separate orbits. Details of the analysis are given in § 3.

Now, if there are n electrons in an atom with coordinates $\mathbf{r}_1 \dots \mathbf{r}_n$, and momenta $\mathbf{p}_1 \dots \mathbf{p}_n$, the total wave function $\psi(\mathbf{r}_1 \dots \mathbf{r}_n)$ which describes the complete electron configuration is related to the total momentum wave function $X(\mathbf{p}_1 \dots \mathbf{p}_n)$ by a formula similar to (1) for a single electron:

$$X(\mathbf{p}_1 \dots \mathbf{p}_n) = (2\pi)^{-3n/2} \int \exp -i\{\mathbf{p}_1 \cdot \mathbf{r}_1 + \dots + \mathbf{p}_n \cdot \mathbf{r}_n\} \Psi(\mathbf{r}_1 \dots) d\mathbf{r}^n. \quad \dots\dots(5)$$

But $\Psi(\mathbf{r}_1 \dots \mathbf{r}_n)$ is the sum of one or more determinants (Slater 1929), and so, on expansion, it may be regarded as the sum of a large number of distinct terms, each of which is the product of n atomic orbitals. Since the coordinates of each electron appear once and once only in each of these terms, the integration (5) for any one term is immediate; the result is simply that which would be obtained by replacing each space atomic wave function by the corresponding momentum wave function. By addition, therefore, it follows that the complete wave function $X(\mathbf{p}_1 \dots \mathbf{p}_n)$ is formally precisely the same as $\Psi(\mathbf{r}_1 \dots \mathbf{r}_n)$ except that each space orbital is replaced by the corresponding momentum orbital calculated as in § 3.

Knowing $X(\mathbf{p}_1 \dots \mathbf{p}_n)$ the calculation of the generalized mean radial distribution function $I(\mathbf{p}_1 \dots \mathbf{p}_n)$ follows at once from an integration completely analogous to (2). All the terms are known, so that I is determined for all atoms.

§ 3. WAVE FUNCTIONS AND FORMULAE

For hydrogen, with only one electron, an exact space wave function $\psi(\mathbf{r})$ is known. In atomic units it is simply $(\pi)^{-1/2} e^{-r}$. For atoms in the second row of the periodic table we have used the analytical wave functions originally proposed

by Morse, Young and Haurwitz (1935) and corrected by the present writers (1944). The resulting mean radial distribution functions have been given in D.C. (p. 193) and need not be reproduced here. For the remaining atoms (11 $Z < 20$) we have used wave functions suggested by Slater (1930); that is to say, the complete wave function for an atom is made up in the usual way (Slater, 1929) as sums of determinants in each of which the electrons are defined by atomic orbitals of the form

$$\psi_{nlm}(\mathbf{r}) = N_r N_\theta N_\phi r^{n^*-1} e^{-cr} P_l^m(\cos \theta) e^{+im\phi}. \quad \dots\dots (6)$$

The N 's are separate normalizing factors for the radial and the two angular terms in ψ , n^* is the effective principal quantum number whose value is given by Slater in terms of the true quantum number, c is a quantity related to the nuclear charge Z and the screening constant s by the formula

$$c = (Z - s)/n^*. \quad \dots\dots (7)$$

In this way all the coefficients in (6) are supposed to be known.

We have also used Slater functions for some of the lighter atoms ($Z < 10$). The results, which we do not need to describe in detail, since they are presumably less accurate than the Morse functions, show that for these atoms there is very little difference between the momentum distributions predicted by the two types of wave function. It seems highly probable, therefore, that the same conclusion will hold for the heavier atoms (11 $Z < 20$) for which the more detailed wave functions are not known.

The transformation from the complete space wave function for all the electrons in an atom to the complete momentum wave function requires, first of all, that we should be able to transform the individual atomic orbitals $\psi(\mathbf{r})$. This transformation from $\psi(\mathbf{r})$ to $\chi(\mathbf{p})$ proceeds as follows. Using (1) and (6):

$$\chi(\mathbf{p}) = (2\pi)^{-3/2} N_r N_\theta N_\phi \int r^{n^*-1} e^{-cr} P_l^m(\cos \theta) e^{+im\phi} e^{i\mathbf{p}\cdot\mathbf{r}} d\mathbf{r}. \quad \dots\dots (8)$$

Now Bauer's expansion of $e^{i\mathbf{p}\cdot\mathbf{r}}$ gives

$$e^{i\mathbf{p}\cdot\mathbf{r}} = \sum_{j=0}^{\infty} (2j+1)(-i)^j P_j(\cos \vartheta) f_j(p, r),$$

where ϑ is the angle between \mathbf{r} and \mathbf{p} , and the functions $f_j(p, r)$ are defined by

$$f_0(p, r) = \sin pr/pr, \quad f_j(p, r) = \sqrt{(\pi/2pr)} J_{j+1/2}(pr). \quad \dots\dots (9)$$

Substitution of these values in (8) allows the integration over the solid angle $d\omega$ [$d\mathbf{r} = d\omega dr$] to be made at once. And if we write p, θ_p, ϕ_p for the polar coordinates of \mathbf{p} , we have

$$\chi(\mathbf{p}) = (-i)^l \sqrt{(2/\pi)} N_r N_\theta N_\phi P_l^m(\cos \theta_p) e^{+im\phi_p} R_l(p), \quad \dots\dots (10)$$

where

$$R_l(p) = \int_0^\infty r^{n^*-1} e^{-cr} f_l(p, r) r^2 dr. \quad \dots\dots (11)$$

The fact that $\chi(\mathbf{p})$ depends on angle in precisely the same way as $\psi(\mathbf{r})$ has already been noted, for hydrogen-like wave functions, by Pauling and Podolsky (1929): our argument shows it to be true for wave functions in any central field.

Using (10) and the definition (2) of $I(p)$ it follows that

$$I(p) = (2/\pi) N_r^2 p^2 R_l(p)^2, \quad \dots\dots (12)$$

where, by a simple integration it may be shown that

$$N_r^2 = (2c)^{2n^*+1} I(2n^*+1). \quad \dots\dots (13)$$

The calculation of $\chi(p)$ and $I(p)$ for any one electron now depends only on a calculation of $R_l(p)$. This must be made separately for s, p, d, \dots electrons for which $l=0, 1, 2, \dots$. If n^* is integral, as it is for the first three rows of the periodic table, the integration (11) yields algebraic functions. These functions are

$$\begin{aligned} I_{1s}(p) &= \frac{32c^5 p^2}{\pi(p^2 + c^2)^4}, & I_{2s}(p) &= \frac{32c^5 p^2 (p^2 - 3c^2)^2}{3\pi(p^2 + c^2)^6}, \\ I_{2p}(p) &= \frac{512c^7 p^4}{3\pi(p^2 + c^2)^6}, & I_{3s}(p) &= \frac{1024c^9 p^2 (p^2 - c^2)^2}{5\pi(p^2 + c^2)^8}, \\ I_{3p}(p) &= \frac{1024c^7 p^4 (5c^2 - p^2)^2}{45\pi(p^2 + c^2)^8}, & I_{3d}(p) &= \frac{4096c^9 p^6}{5\pi(p^2 + c^2)^6}. \end{aligned} \quad \dots\dots (14)$$

Care must be exercised in this and later formulae to remember, from (7), that c has a different numerical value for each orbit in each atom, and the appropriate value must always be used in any particular case. Different values of c in any single formula, however, merely correspond to changes in scale of p .

For electrons of higher quantum number than those in (14), n^* is not integral, and the evaluation of $R_l(p)$ is more clumsy. It may be systematized most simply in terms of two new sets of functions $C(n)$ and $S(n)$, defined by

$$C(n) = \int_0^1 r^n e^{-pr} \cos pr \, dr, \quad S(n) = \int_0^1 r^n e^{-pr} \sin pr \, dr. \quad \dots\dots (15)$$

$C(n)$ and $S(n)$ are functions of c and p , as well as of n , and if we put

$$p = c \tan \alpha, \quad (0 < \alpha < \pi/2), \quad \dots\dots (16)$$

we may write

$$\begin{aligned} C(n) &= \Gamma(n+1) \cos(n+1)\alpha \cos^{n+1}\alpha, \\ S(n) &= \Gamma(n+1) \sin(n+1)\alpha \cos^{n+1}\alpha. \end{aligned} \quad \dots\dots (17)$$

In terms of these functions it may be shown that

$$\begin{aligned} pR_0(p) &= S(n^*), \\ p^2R_1(p) &= S(n^*-1) - pC(n^*), \\ p^3R_2(p) &= 3S(n^*-2) - 3pC(n^*-1) - p^2S(n^*), \\ p^4R_3(p) &= 15S(n^*-3) - 15pC(n^*-2) - 6p^2S(n^*-1) + p^3C(n^*). \end{aligned} \quad \dots\dots (18)$$

From these relations, with (10) and (12), both $\chi(p)$ and $I(p)$ may be calculated for all electron orbits of s, p, d and f type. The simple results in (14) are, of course, merely particular cases of (18) and (12) in which n^* is integral.

§ 4. THE CALCULATIONS AND RESULTS

As determined above $I(p_1 \dots p_n)$ is a function of the magnitudes $p_1 \dots p_n$. The simplest way of expressing it pictorially is to integrate over all values of $p_2 \dots p_n$ from 0 to ∞ , obtaining a function $I(p_1)$, which we may call the mean

radial momentum density for electron 1. Now, by the very nature of the determinantal wave functions used (which results from the fact that all electrons are identical), $I(p_1) = I(p_2) = \dots = I(p)$, say. So we may regard $I(p)$, which is normalized to unity, as measuring the velocity distribution in the atom, just as the more familiar density $\rho(r)$ measures the space distribution. It is this quantity $I(p)$ which we have calculated.

If all the original atomic orbitals (6) are orthogonal, it follows that $I(p)$ is simply the sum of contributions such as (12) from each electron present. In our calculations for the atoms of the first two rows of the periodic table, this condition is accurately fulfilled. But for the others it is not, and although orbitals with different l values are rigorously orthogonal on account of the surface harmonic terms in (6), there is a small non-orthogonality between $1s, 2s, 3s, 4s$, and between $2p, 3p, 4p$ etc. Calculations (D.C., p. 198) for Li and C show that the result of neglecting this non-orthogonality (which is rather clumsy to include) is to change the width of the Compton profile (§5) by less than 1%. Now it would be presumptuous to claim an accuracy as good as this for the initial wave functions: accordingly, for the heavier elements for which $11 < Z < 20$, we have made our calculations of $I(p)$ on the assumption that all the atomic orbitals may be regarded as orthogonal. The curves of $I(p)$, therefore, which we show in figures 1 and 2 are simply sums of curves corresponding to (14) and (18), suitably weighted according to the numbers of each type of electron present. In every case, of course,

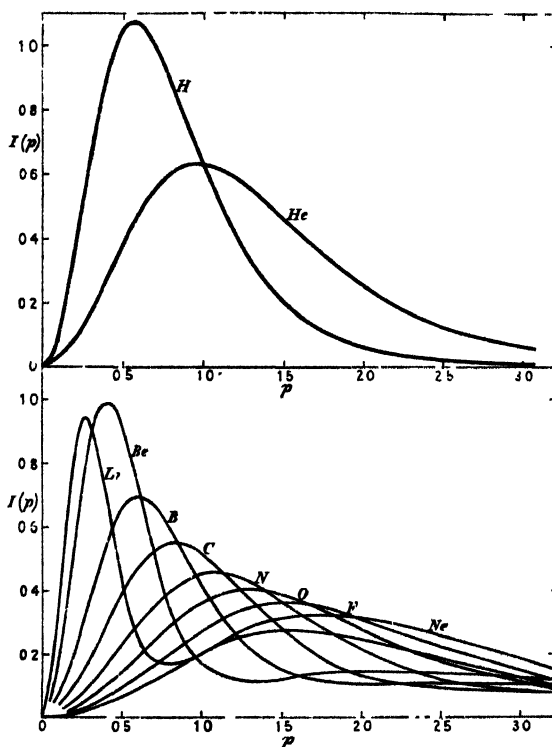


Figure 1.

$$\int_0^{\infty} I(p) dp = 1.$$

Certain deductions from these curves suggest themselves at once:—

(a) First of all, there is considerable resemblance between the $I(p)$ curves in figures 1 and 2 and corresponding electron density curves. For example, just as in the case of space wave functions, there are momentum shells within an atom. But the behaviour of these shells is exactly opposite to that with space functions. For as we proceed along any row of the periodic table, figures 1 and 2 show that the shells expand (instead of contracting), and when we start a new shell, it appears at the inside (instead of the outside) of the ones already there. The writers have never seen this expressly stated before, but it is implicit in the

Fourier relationship of the two types of wave function; for it is a known result that a function and its transform cannot both be made dense in any given region: and this, of course, is merely the Heisenberg Uncertainty Principle in its fundamental form.

An alternative physically significant interpretation of the behaviour of each new shell is that the electrons in the new shell are at first loosely bound, and therefore have, on the average, small momenta. As the shell is completed, the binding increases, so that the mean momentum increases also.

The steady drift to larger values of momentum with increasing nuclear charge is well illustrated by the behaviour of the $1s$ electrons. According to (14) the peak of the $1s$ curve lies at $p = c/\sqrt{3}$. But c varies from 1 for hydrogen to 18.7 for potassium, so that the peak moves steadily from 0.58 A.U. to 10.8 A.U. At the same time the height of the peak falls, for the height per electron is $27/8\pi c$. When the contributions are weighted according to the number of electrons present, this means that the maximum contribution from the $1s$ electrons falls from 1.05 for hydrogen to 0.01 for potassium. After this stage we may say that the $1s$ electrons are scarcely relevant to the density function except for very large values of p . This is quite different from the behaviour of the space wave functions, where $\psi(1s)$ crowds steadily in towards the nucleus, and may eventually be regarded as completely screening the nucleus.

(b) We also notice that s -electrons give uniformly narrower, or more compact, $I(p)$ curves than the corresponding p -electrons. This means that the alkali atoms for which the outer shell is simply s , and, even more so, the atoms of the next column of the periodic table, for which the outer shell is s^2 , show a pronounced sharp peak in the $I(p)$ curve. But progressive addition of p -electrons smoothes this out so that, for example, whereas in Li there is a very clear separation between the $2s$ and the $1s$ contributions, by the time that we have reached

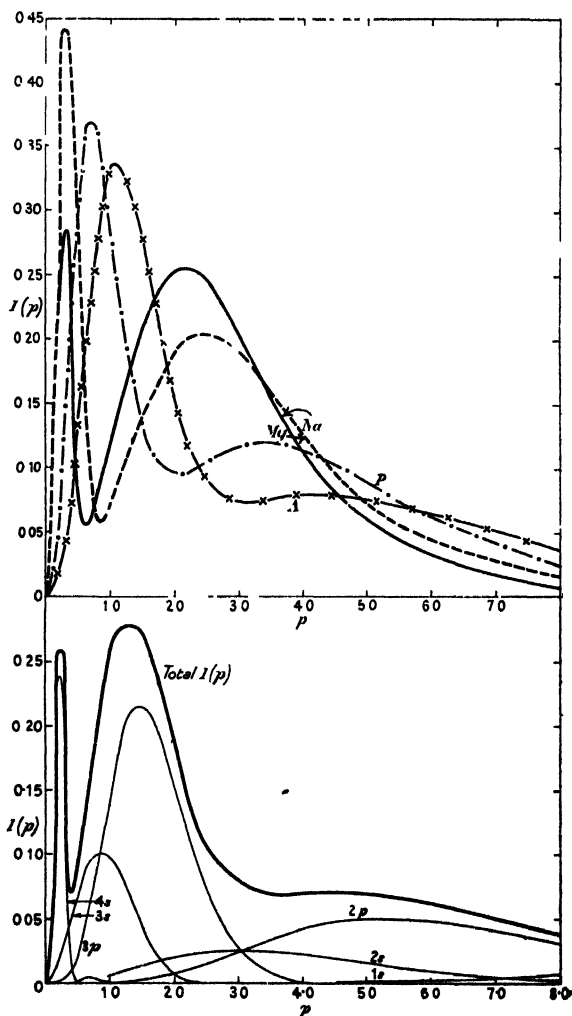


Figure 2

carbon the distinction is obliterated and one single wider $I(p)$ curve results. There is, in fact, a gradual transition as we proceed along the second row of the periodic table from Li to Ne, and also along the third row from Na to Ar. In each case, at the beginning of the next rows (Na, K) a new inner peak of great sharpness appears.

(c) In the particular case of potassium, shown at the bottom of figure 2, we have drawn both the total $I(p)$ curve and the components, with correct weight, that comprise it. In the total curve, one can clearly distinguish the inner peak due to the $4s$ contribution, followed by a second peak arising from $3s$ and $3p$ electrons, next comes a still wider peak from the $2s$ and $2p$ electrons and there is a very low flat peak at large distances arising from the $1s$ electrons. Similar discussions could be given for the other atoms, but this sufficiently illustrates the way in which the different electrons contribute to $I(p)$.

A quantity of some importance is the mean momentum \bar{p} . According to what we have said about $I(p)$ and the formula (4), this is the sum of suitably weighted contributions $\bar{p}_1, \bar{p}_2, \dots$. Explicit calculation shows that

$$\bar{p}_1 = 8c/3\pi, \quad \bar{p}_2 = 8c/5\pi, \quad \bar{p}_3 = 128c/105\pi, \quad \bar{p}_{2p} = 128c/45\pi, \quad \bar{p}_{3p} = 9472c/4725\pi$$

.....(19)

Values of \bar{p} for other orbitals may be obtained most easily by numerical integration. Figure 3 shows \bar{p} plotted against the atomic number Z . Along each row of

the periodic table the variation of \bar{p} is almost linear, but the slope of each successive shell is less than the one preceding, and there is only a small change in \bar{p} on converting a rare gas to an alkali metal atom. All this is rather as one would have expected because \bar{p} must be related to the total electronic energy E in some such way as $\bar{p} \sim \sqrt{(2E/Z)}$, all in atomic units. Along each row of the periodic table there is a steady increase in

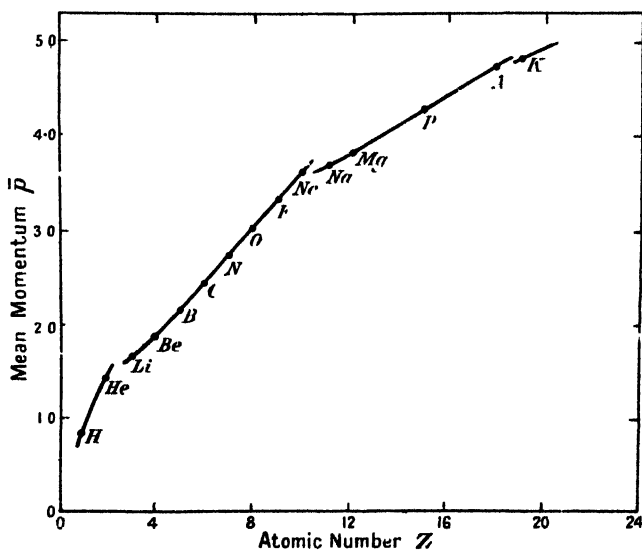


Figure 3.

E/Z , but on account of the poor binding of the one valence electron in the alkali atom, there is only a small change in E on starting a new shell.

§ 5. COMPTON PROFILES

It was shown in D.C. §2 that a knowledge of $I(p)$ enables us to calculate the shape of the Compton profile. All Compton profiles for a given atom (i.e. all incident wavelengths and all angles of scatter) are included, by suitable change

of scale, in a single formula, where the intensity J at "reduced" distance q from the centre of the displaced band, is given by

$$J(q) = \frac{1}{2} \int_0^q \frac{I(p) dp}{p}. \quad (20)$$

Full particulars are given in D.C. and do not need to be reproduced here. In that paper $J(q)$ curves were plotted for all atoms up to Ne. Our present calculations allow us to extend this work up to K. Five representative examples, each giving one-half of the complete profile, are shown in figure 4. From these the general trend of the curves is apparent.

These curves are normalized in each case so that $J(0) = 0$. This, while making the areas under the curves different, greatly facilitates comparison of different profiles. In general, the behaviour of these curves agrees closely with that found for the lighter atoms in D.C., that is, s -electrons give narrow profiles and p -electrons wider ones.

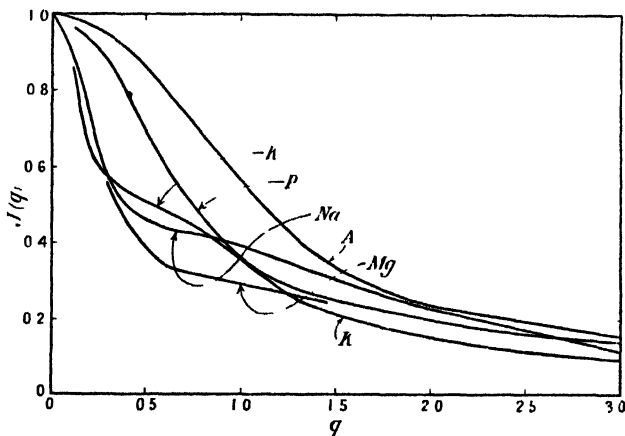


Figure 4.

Thus the sharpest curve in figure 4 is for Mg, whose outer structure is $(3s)^2$, and the widest is for A, where the outer structure is $(3s)^2(3p)^6$. In the cases of Na and Mg there is an appreciable "lag" at the bottom of the curve, similar to that found before for Li and Be; this might easily cause trouble when separating the contributions to the full experimental profile that arise from satellite lines (e.g. the $K\alpha_1$ and $K\alpha_2$ doublet) in the incident radiation.

Finally, we give below, in tabular form, the theoretical widths Δq at half-maximum value, together with the resulting Δl values, in x-ray units, for back-scattering ($\theta = 180^\circ$) of Mo $K\alpha$ radiation. This table, which may be compared exactly with table 1 in D.C., p. 195, shows again a steady increase, except for Na to Mg, along this, the third row of the periodic table, and the smaller value with which each new row begins.

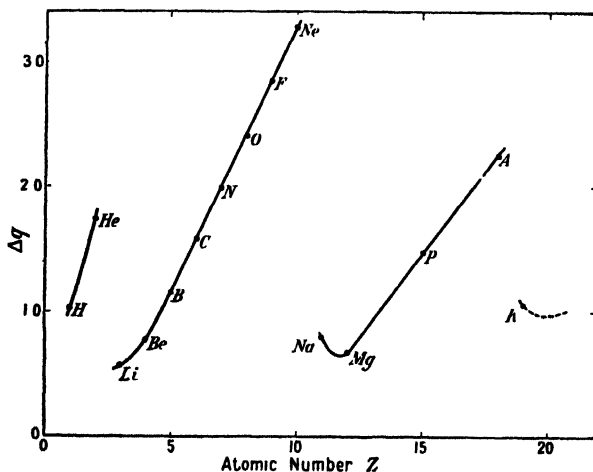


Figure 5

Atom	Ne	Na	Mg	P	A	K
Δq	3.27	0.80	0.68	1.48	2.24	1.06
$\Delta/(x.u.)$ for $\text{MoK}\alpha$ $\theta = 180^\circ$ }	34	8.3	7.1	15	23	11

Thus in the case of the Compton profile, the changes to be expected between one row of the periodic table and the next row are less than between the elements of either row alone. This is shown very clearly in figure 5, which collects together the Δq values for the atoms both of this paper and the preceding one.

REFERENCES

- DIRAC, 1947, *The Principles of Quantum Mechanics*, Chapters 3 and 4 (Oxford: University Press. Third edition).
 DU MOND, 1933, *Rev. Mod. Phys.*, **5**, 1.
 DUNCANSON, 1943, *Proc. Camb. Phil. Soc.*, **39**, 189.
 DUNCANSON and COULSON, 1944, *Proc. Roy. Soc. Edinb.*, A, **62**, 37.
 DUNCANSON and COULSON, 1945, *Proc. Phys. Soc.*, **57**, 190.
 HUGHES and MANN, 1938, *Phys. Rev.*, **53**, 50.
 MORSE, YOUNG, and HAURWITZ, 1935, *Phys. Rev.*, **48**, 948.
 PAULING and PODOLSKY, 1929, *Phys. Rev.*, **34**, 109.
 SLATER, 1929, *Phys. Rev.*, **34**, 1293.
 SLATER, 1930, *Phys. Rev.*, **36**, 57.

Some Experiments with Adjustable Geiger-Müller Counters

BY MOHAMMED CHAUDHRI, Ph.D.(Cantab), M.Sc.(Alig.)*

AND A. G. FENTON, B.Sc.†

Physics Department, University, Birmingham

* Of the Muslim University, Aligarh, India. † On leave from the University of Tasmania, Australia.

MS. received 12 June 1947

ABSTRACT. Special counters are described in which it is possible to alter the effective length as well as the material and diameter of the anode without opening them. This is a distinct advantage in that the effects of these factors on the counter characteristics can be studied without complication due to changes in the gas filling and the nature of the cathode surface brought about by opening the counters or by making separate counters with anodes of different diameters.

It was observed that, while in a given counter the anode material and the nature of its surface have little or nothing to do with the characteristics, there is an optimum anode diameter for which the best plateaux are obtained, a factor of obvious practical importance in counter design.

The photosensitivity to visible light and longer wavelengths caused by high counting rates or by "continuous" discharges was shown to be due to some surface phenomenon on the cathode.

It has been shown in a simple manner that the discharge is not localized, but spreads along the length of the counter.

It was observed that the plateau length and slope improve at first, but then deteriorate when the effective length of the counter is reduced by shielding the anode wire at both ends with glass capillary tubes. With metal capillaries, however, the plateau characteristics are best when the full length of the cathode is used.

§ 1. INTRODUCTION

MUCH work has been done on Geiger counters of the self quenching type (Korff 1946) but very few systematic investigations have been carried out which yield reliable data as to the relative importance of the various components of a counter during its operation. Moreover, there are still a number of factors influencing the plateau characteristics of counters which have not been given due consideration by previous workers.

It is the purpose of the present paper to study the effects of the following factors on the plateau characteristics:—

- (1) The material of the anode wire.
- (2) The diameter of the anode wire.
- (3) The shielding of the anode wire.
- (4) The cathode surface.

It has also been found during the course of these experiments that the cathode surface can be made photosensitive to wavelengths up to about 1 micron, and that the degree of the photosensitivity and its duration can be controlled within limits.

§ 2. APPARATUS

The experiments were performed with four counters having nickel cathodes and one counter with an aluminium cathode. The nickel cathodes were made from sheet nickel 0.1 mm. thick formed into cylinders 2.2 cm. in diameter and 10 cm. long, spot-welded along the seam. The aluminium cathode was a tube 0.1 mm. thick, 10 cm. in length and 2.0 cm. diameter. The cathodes were cleaned with nitric acid and were rinsed thoroughly with water and dried before assembly into the counters.

One of the counters with a nickel cathode has an anode made from a number of wires of different materials and of different gauges spot-welded together end to end. In the discussion which follows we shall refer to this as counter number 1 and we give below a detailed description of it and of the other counters.

The envelope of the counter is a Pyrex glass tube 60 cm. long and 2.5 cm. in diameter. The nickel cathode C, figure 1, is closely fitted into the glass envelope. Electrical connection of C with the external circuit is made by means of a nickel strip spot-welded to the cathode at one end and to the Pyrex-tungsten seal S at the other.

The anode A consists of wires of gold, silver, platinum, copper, molybdenum, tantalum and tungsten, ranging in diameter from 0.065 mm. to 0.25 mm. Any one of these wires may be set within the cathode by rotating the vacuum-tight ground-glass joints E which support the glass tubes F on which the anode is wound. The extreme ends of the anode are spot-welded to 1 mm. nickel wires which run through the glass tubes F and are held in position by the Pyrex-tungsten seals J, so that electrical connection to the anode may be made from outside the tube. The anode passes through the 2 mm. bore Pyrex capillary tubes D, each 15 cm. in length and 5 mm. in external diameter. The outer ends of the capillaries are tightly fitted into the soft iron cylinders B, each 3 cm. long and 2.3 cm. in diameter, which slide smoothly within the outer glass envelope. The weight of the glass capillaries is supported by two glass tubes G with a bore just greater than the external diameter of the capillaries, internally sealed into the envelope. The

ground-glass joints K carry thick leads L through which current may be passed to heat the anode wire when desired. The capillaries may be set in any desired position inside the cathode by moving the soft-iron cylinders from outside with the help of a magnet.

Counter 2 has a nickel cathode similar to that of number 1 but has a fixed anode of 0.2 mm. diameter tungsten wire sealed at the ends of the glass envelope, which is 45 cm. long and 2.5 cm. in diameter. The arrangement of the glass capillaries is the same as in counter 1. Counter 3 has the same construction as 2 but brass capillaries of the same dimensions as in 2 are used instead of glass. In counters 4 and 5 a single adjustable Pyrex glass capillary (15 cm. long and 5 mm. in diameter and 2 mm. bore), sliding over the 0.2 mm. tungsten anode wire is used, the motion of the capillary being controlled mechanically through a ground-glass joint. The cathodes are of the same dimensions but one is of nickel, the other of aluminium.

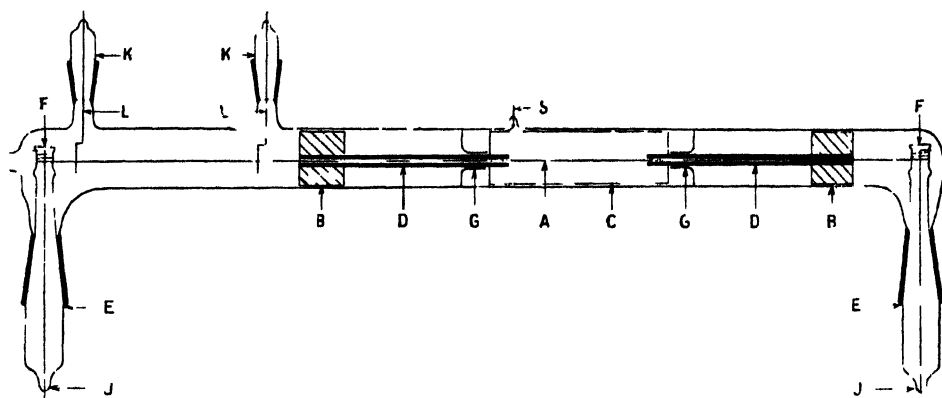


Figure 1

The counters were evacuated with a mercury diffusion pump and were out-gassed by warming and by letting the pumps run continuously for several hours. The counters were tested for vacuum tightness by allowing them to stand under vacuum for several days.

The gas filling used was an argon-alcohol mixture which was allowed to stand in a flask for over 24 hours to mix thoroughly before being admitted to the counters.

Counters 1, 2 and 3 were filled with the same argon-alcohol mixture containing 23% alcohol, to a pressure of 10 cm. Hg. Counter 2 had been used previously with fillings of different gases but 1 and 3 were new. Counters 4 and 5 were filled with an argon-alcohol mixture containing 10% alcohol, to 10 cm. total pressure. Once filled to the desired pressure the counters were not opened or refilled during the experiments described below.

No wax joint was used in any part of the counters or apparatus and the grease used for the taps and ground-glass joints was apiezon M.

A scale-of-100 recorder was used to record the counts, and the potentials for the counters were obtained from a stabilized H.T. set giving up to 1650 volts.

§ 3. RESULTS

Effective length and plateau characteristics

Preliminary experiments with counters 4 and 5 indicated that the distance between the ends of the capillary tubes within the cathode determines the effective length of the counter. Moreover, it was observed that the plateau slope improved when the effective anode length was reduced by sliding the capillary into the cathode. Detailed measurements were

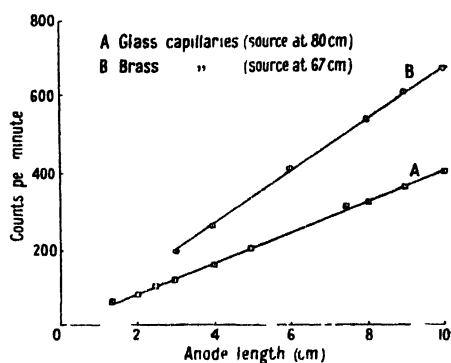


Figure 2.

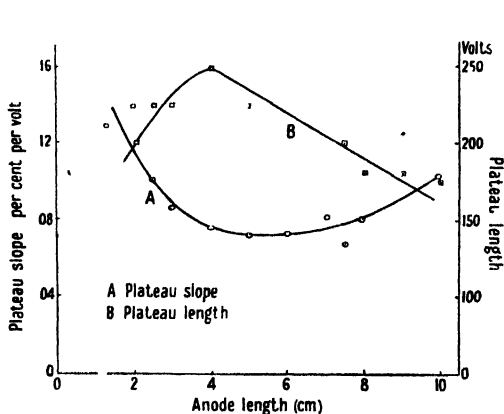


Figure 3

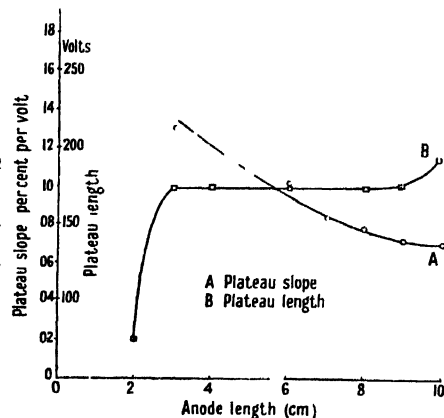


Figure 4

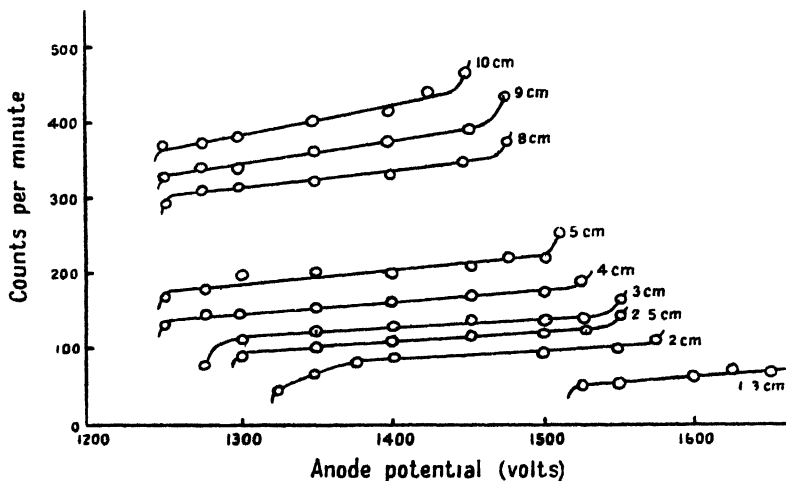


Figure 5

made with counters 2 and 3, in which both capillaries are adjustable. Figure 2 shows the relation connecting the length of the anode wire exposed to the cathode between the ends of the capillaries and the counting rate at a working potential of 100 volts above the threshold (i.e. about the middle of the plateau) when a gamma-ray source is kept at a fixed distance from each counter. The relation between the effective anode length and the slope of the plateau for counter 2 is shown in figure 3, curve A, while curve B shows the relation between the effective anode length and the length of the plateau. The curves indicate that as the anode length is reduced from the maximum of 10 cm. the plateau slope and length improve at first but later deteriorate. Similar curves in figure 4 show that for counter 3 the plateau is best at 10 cm. anode length and begins to deteriorate as the metal capillaries are slipped into the cathode. Figure 5 shows the characteristic curves of counter 2 for different lengths of the anode. The variation of the threshold voltage with the effective anode length is shown in curves A and B, figure 6, for counters 2 and 3 respectively. As the effective length is reduced below 2 cm., the threshold voltage increases rapidly until at about 1 cm. there are no counts even at 1600 volts.

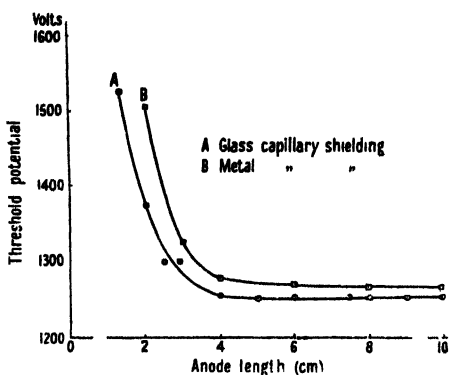


Figure 6

§4 INFLUENCE OF THE GAUGE AND MATERIAL OF THE ANODE ON PLATEAU CHARACTERISTICS

The plateau characteristics of counter 1 were obtained using anodes of different materials and of different diameters. As the anode could be changed by rotating the ground-glass joints, other conditions could be kept constant during the measurements. A 4 cm. anode length at the centre was used throughout these tests, as earlier work had shown that good plateaux are obtained with this length. Figure 7, curve A, shows the relation between the plateau slope and the wire diameter, while curve B shows how the length of the plateau varies with the wire diameter. It was found that anode wires of a given diameter give the same plateau characteristics although made from different materials.

Figure 7, curve C, shows that the threshold voltage decreases when the diameter of the anode is reduced.

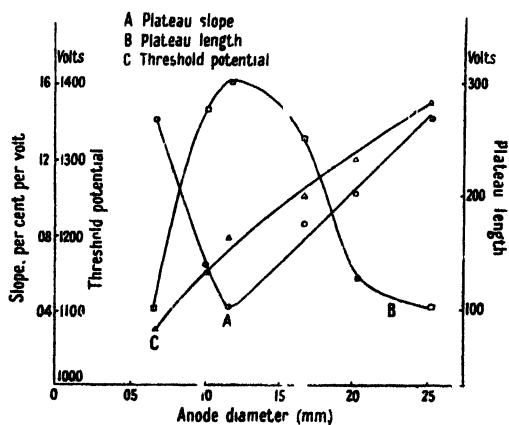


Figure 7

§ 5 PHOTSENSITIVITY OF THE COUNTERS

Early experiments with counter 4 showed that after passing a "continuous" discharge* through the counter (by setting the voltage a few hundred volts above the plateau region) the background counting rate immediately afterwards was of the order of 10000 counts per minute, but after 24 hours it had fallen to 200 per minute, the normal background being about 50 counts per minute and the plateau characteristics were slightly better than before the discharge. When counter 5, which has an aluminium cathode, was new, it had plateau characteristics as good as counter 4, but after a continuous discharge it required a considerably longer time to recover and was thereafter unreliable for counting rates of over a few hundred counts per minute. In subsequent measurements, which were carried out in greater detail with other counters, it was found that these effects were primarily due to changes in the behaviour of the cathode surfaces, produced by the discharge.

Counter 2, with which a few million counts had already been recorded over a period of some weeks without making it photosensitive, was run at a high counting rate at an operating voltage about 100 volts above the threshold. In a particular case, an anode length of 3 cm. was used and gamma rays were counted at a rate of about 6000 counts per minute. Background counts taken with the room lights on, following doses of varying duration, showed that a detectable photosensitivity was built up after counting at this rate for 10 minutes. After a few minutes, the background counting rate fell to within probable error of the normal value. The degree of photosensitivity and the period for which it lasts depend upon the dose given.

The photosensitivity caused by passing a continuous discharge was studied in some detail with counter 1 in order to find out what part of the counter becomes photosensitive. The procedure adopted was as follows:

The capillaries in counter 1 were set so that a 3 cm. length of the anode, starting 1 cm. from one end of the cathode, was used during the discharge. Since the cathode is 10 cm. in length, a fresh portion of it is used when the capillaries are adjusted for a similar 3 cm. length at the other end of the counter. After passing a discharge for 5 minutes, the background counting rate with the room lights on was about 20 times the normal value, but the rate on moving the capillaries to the other end of the counter was normal. Thus it appears that this portion of the counter had not been affected at all by the discharge, and that the gas filling retained its quenching characteristics. Next, the portion of the anode which had been used during the discharge, was wound along until it became the anode for the other end of the counter. Again, a normal counting rate was found, while the photosensitivity persisted at the end in which the discharge had been passed. This means that the portion of the cathode which was used during the discharge had become photosensitive. The sensitivity is greater and persists for a longer time when the duration of the discharge is increased.

After a discharge had been passed in this counter for 1 hour, the photosensitivity decayed as shown in figure 8, curve A. The sensitivity decreased rapidly at first and then more slowly. In one counter appreciable photosensitivity remained after it had been standing unused for a period of several months. The data of curve A, figure 8, are shown on a double logarithmic plot in curve B.

* When the potential is set above the Geiger region, the discharge which passes is not really continuous, but consists of discrete pulses in rapid succession.

Two days after a continuous discharge had been passed for 1 hour in counter 1, the spectral sensitivity was tested qualitatively with a set of Wratten filters and a tungsten-filament lamp, as after this time the decay of the photo-sensitivity was too slow to alter appreciably during the counting intervals. It was found that, under these conditions, the maximum sensitivity in the visible region was at the short wavelength end of the spectrum and there was no detectable sensitivity above about 5500 Å. Immediately after a continuous discharge however, the counter was always sensitive to longer wavelengths than this, but measurement was difficult because of the rapid decay rate.

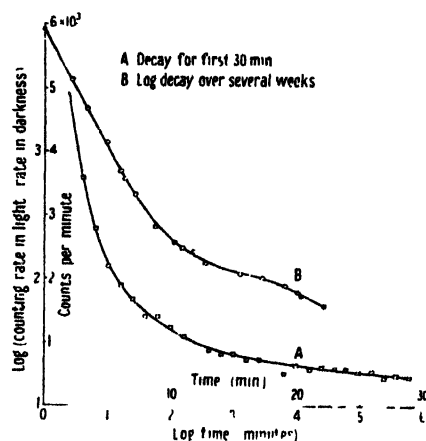


Figure 8

§ 6. EFFECT OF A CONTINUOUS DISCHARGE ON THE PLATEAU CHARACTERISTICS

The plateau characteristics were obtained for two 3 cm. effective lengths of counter 2, each set at a distance of 1 cm. from the ends of the cathode. A discharge was then passed for 1 hour in one of these sections, and immediately afterwards it was observed that the slope and length of the plateau as well as the background in the other section with room lights on, were, within probable error, the same as originally. On the other hand, the background for that section of the counter in which the discharge had been passed was too high for the recording apparatus to follow. In darkness, however, the background for this end of the counter was the same as before the discharge and equal to that of the other end in light. Though the background of this part of the counter was too high in light for several days for the recorder to follow, its plateau characteristics taken with room light off were as good as those obtained in light before the discharge.

§ 7. EFFECT OF A CONTINUOUS DISCHARGE ON THE BACKGROUND IN DARKNESS

During the experiments with counter 1 it was observed that the background counting rate with the tube in darkness was higher than normal immediately after the discharge had been passed in it for an hour, but returned to the normal value within 48 hours. This phenomenon was studied in detail with counter 4, in which the effect was very pronounced. Figure 9, curve A, shows the decay of the background counting rate in

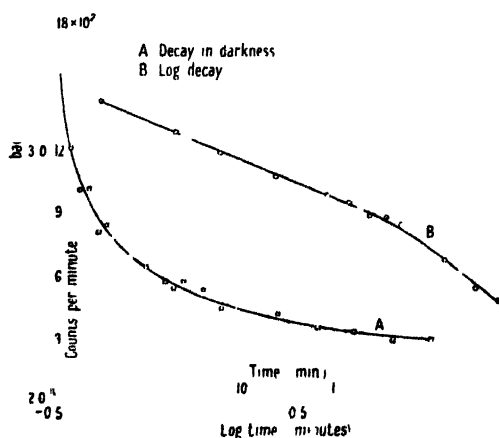


Figure 9

darkness during the 20 minutes interval immediately after a continuous discharge of 30 minutes. It was found that this effect disappeared altogether after a number of discharges of short duration had been passed in the counter, although it could still be made highly photosensitive. This was found also to be the case with other counters.

§8. DISCUSSION OF RESULTS

During the experiments the pulses from counter 2 were examined with a cathode-ray oscillograph and it was observed that for a given operating potential the pulse amplitude increased almost linearly with the effective anode length, which is consistent with the assumption that the discharge is not localized but spreads along the length of the wire. This demonstrates in a simple way the conclusion arrived at previously by other workers (Wilkening and Kanne 1942, Stever 1942) and studied recently by Hill and Dunworth (1946), Craggs and Jaffe (1947), and Wantuch (1947).

The linear relation (figure 2) between the separation of the ends of the capillaries and the counting rate shows that the efficiency of the counter is unaltered by reducing its length and also that for practical purposes we are justified in referring to the separation of the capillaries as the effective anode length of the counter. At very short lengths, however, the distortion of the fields near the ends of the capillaries would be expected to have a noticeable effect and this appears to be the case, since the counter with glass capillaries was not sensitive to gamma rays at an anode length less than 0.8 cm. with a working potential of 1600 volts; while the metal capillary counter at the same operating voltage would not respond below 1.25 cm. anode length.

The slope of the plateau in self-quenching counters has been attributed by Korff (1944) to the occurrence of multiple pulses which, with a recording circuit of high resolution, give rise to spurious counts. On this view the more ions there are per pulse the greater is the probability of a multiple pulse occurring. Therefore, when the pulse size is reduced by reducing the anode length of a counter, one would expect to obtain a longer and flatter plateau. This would account for the initial improvement in the plateau characteristics for the counter with glass capillaries (figure 3) and for the increase in the upper limit of the plateau as the anode length is reduced (figure 5). Figure 5 shows that as the anode length is reduced below the 4 cm., the threshold moves to higher voltages more rapidly than the upper limit of the plateau, resulting in a reduction of the plateau length. This increase of threshold potential with decreasing anode length (figure 6) may be attributed at least partly to the sensitivity of the recording apparatus which does not respond to pulses of less than a certain minimum amplitude. Observation with a cathode-ray oscillograph during the experiments showed that this minimum was reached when the anode length was reduced to about 4 cm. with a working potential of 1250 volts. At shorter lengths than 4 cm. the plateau did not begin until the pulse size was increased by using higher operating potentials. The increase in slope of the plateau for short anode lengths suggests that spurious counts due to factors other than the number of positive ions per pulse were recorded. These may be due to point discharges from irregularities on the surfaces of the electrodes or to distortion of the field at the ends of the capillaries,

both of which effects would be more noticeable at the higher potentials used. Furthermore, at higher potentials, the average energy of the positive ions being greater, there is a greater chance for some of them to reach the surface of the cathode and liberate secondary electrons, thus giving rise to spurious counts. In the counter with the brass capillaries, the best plateau was obtained at 10 cm. anode length, suggesting that the expected improvement in the plateau during the initial stages on reducing the anode length was masked by effects caused by the presence of the metal capillaries.

In order to explain the high background in darkness, observed after a continuous discharge (figure 9), one must look for processes occurring on the cathode surface in which electrons or negative ions are emitted. One possibility is that during the discharge active centres are formed on the cathode surface, and these undergo chemical change with the emission of electrons or ions. Also, if during this chemical change, positive ions are formed on an insulating layer, e.g. oxide, on the cathode surface, these may give rise to *field emission* electrons. On the other hand, the photosensitivity which is observed when the cathodes are in an activated condition after a high counting rate or a continuous discharge suggests that centres of low work-function are formed. With a sufficiently low work-function, and under the influence of the electric field at the cathode, an observable number of thermionic electrons would be able to escape at room temperature, thus accounting for the high background in darkness. For example, calculation shows that, if we regard the activated cathode as similar to an oxide coated thermionic emitter for which A in the emission equation $i = AT^2 e^{-\phi/kT}$ is 0.03 amp. per sq. cm. and ϕ is 1.2 volts, the current at 300° K. is about 10 electrons per sq. cm. per second. This view is supported by the observation that the counter was sensitive to wavelengths up to about 10000 Å. when the high dark background was also present. Furthermore, the counter was sensitive to the long wavelengths only for a short period immediately after a discharge, but sensitivity at the low wavelength end of the visible spectrum remained for a long time, suggesting that the work function of the centres was low at first but increased with time. It is also worth noting that the counting rate in darkness increased when the counter was slightly warmed with hot air from a hair dryer, an effect quite distinct from the photosensitivity at long wavelengths.

High background rates in darkness after high counting rates have been reported by several workers. Lauterjung and Neuert (1944) found the effect most marked for cathodes of low work-function such as Mg (photoelectric work function 2.4 volts), whereas in the present work the effect was observed with nickel cathodes with a work function of about 5 volts. Christoph (1935) reported that in photon counters with Cd or K cathodes the background immediately after a glow discharge was much higher than normal. He also observed an increased photosensitivity after the glow discharge. Lauterjung and Neuert obtained linear decay curves using double logarithmic plots, whereas in our case the curve is linear at first and later changes slope (figure 9, B).

The upper portion of the curve of figure 8, B, representing the first 30 minutes decay in light after a continuous discharge for 1 hour is linear and has a slope of unity which suggests that the natural decay curve follows a rectangular hyperbola law during that period but afterwards the decay becomes more complicated.

After 4 weeks the background in light was about twice the dark background and the decay curve became linear again, indicating that if it were to continue in this way, a measurable sensitivity should remain after 10^6 minutes (i.e. almost 2 years). Actually, one of the counters was found to be photosensitive 4 months after a continuous discharge of 3 hours duration had been passed in it.

Spatz (1943) reported that counters with silver cathodes became photo-sensitive to visible light after long use, and showed how the plateau was destroyed by running the counters for some time at a high counting rate, but gradually recovered again over a period of several days. In view of the present work, we may suppose that the results obtained by Spatz were probably due to the effects observed by us and discussed in this paper.

Figure 7 shows that for a particular counter there is an optimum anode diameter which results in a long, flat plateau, a point which is of great practical importance.

It was invariably found during the work that the threshold occurred at a slightly lower potential and the plateau was longer when the counters were new or operated after a rest period of a few hours. Some minutes of operation, however, brought the threshold to its higher steady value and the plateau characteristics thereafter were reproducible so long as care was taken not to allow the counter to go into a continuous discharge. Observation with an oscillograph showed that after a rest period the pulses were larger, and this may explain the lower thresholds.

It was also observed that the degree of oxidation of the anode wire does not influence to any appreciable extent the characteristics of a counter, because just as good plateaux were obtained using a heavily oxidized copper anode as with a gold wire of the same diameter. Moreover, no difference was found when the anode wire was slackened so that it was about 2 to 3 mm. off the axis at the centre of the counter.

It is generally believed that after prolonged use the plateau characteristics of a counter deteriorate as a result of the dissociation of the quenching agent. Our results suggest, however, that this may be more closely connected with the changes on the cathode surface than with the breakdown of the filling.

It is hoped to continue the investigation of some of the aspects discussed above.

§ 9. ACKNOWLEDGMENTS

It is a great pleasure to record our sincere thanks to Professor M. L. Oliphant, F.R.S., for providing the facilities for this work, and for his continued interest and guidance during its performance. One of us (R.M.C.) owes thanks to him for the award of a Nuffield Research Fellowship, and the other (A.G.F.) is grateful to Messrs. Cadbury-Fry-Pascall Ltd., Claremont, Tasmania, for the Overseas Fellowship which enabled him to take part in this research. Thanks are due also to our respective Universities for granting us leave of absence.

We wish to thank Mr. H. J. Morris for making the glass parts of the counters and assisting us in their assembly.

REFERENCES

- CHRISTOPH, W., 1935, *Ann. Phys. (Chim.)*, **23**, 747.
 CRAGGS, J. D., and JAFFE, A. A., 1947, *Nature, Lond.*, **159**, 369.
 HILL, J. M., and DUNWORTH, J. V., 1946, *Nature, Lond.*, **158**, 833.
 KORFF, S. A., 1944, *Phys. Rev.*, **65**, 274.
 KORFF, S. A., 1946, *Electron and Nuclear Counters* (Van Nostrand). (The book contains an excellent bibliography on counters.)
 LAI TERJUNG, K., and NEUERT, H., 1944, *Z. Phys.*, **122**, 266.
 STEVER, H. G., 1942, *Phys. Rev.*, **61**, 38.
 WANTUCH, E., 1947, *Phys. Rev.*, **71**, 646.
 WILKENING, M. H., and KANNE, W. R., 1942, *Phys. Rev.*, **62**, 534.

The Calibration of Hydrophones and Crystal Transceivers

By N. F. ASTBURY

formerly of H.M. Underwater Detection Establishment

MS. received 1 May 1947

ABSTRACT. A scheme of measurements is discussed from which absolute determinations of sound field pressure can be made. The relation between the scheme and the reciprocity method is discussed. It is shown that the measurements also yield the axial pressure and "projection efficiency" of a transceiver, the "projection efficiency" being defined as the ratio of the actual axial intensity to the axial intensity as it would be if the transceiver converted the whole energy absorbed into sound by vibrating as a simple piston. It is suggested, without proof, that the upper limit of the projection efficiency is unity, and the relation between this quantity and the efficiency defined as the ratio of radiated acoustic power to consumed electrical power is shown to depend on a directivity coefficient and on the effective area of the radiator face. It is pointed out that neither of these quantities is readily determinable and that therefore the conversion efficiency of the transceiver, and hence the acoustic power radiated, cannot be obtained.

§ 1. INTRODUCTION

THE determination of the axial pressure, efficiency and directional properties of crystal transceivers used in submarine acoustic work presents many problems which, in the urgency of immediate need, have tended to remain without complete solution. The present paper is a report of a theoretical investigation into the possibilities of making exact measurements on the quantities specified for crystal transceivers.*

An obvious initial requirement is a device for the absolute measurement of sound field pressure. Such devices are provided by the Rayleigh disc and the radiometer disc (which latter measures energy density in the wave) and present considerable technical problems of their own. On the other hand, it is known that by the use of the reciprocity theorem it is possible to measure sound pressures without any "absolute" device. The method described in this paper is closely related to that based on the reciprocity theorem (Foldy and Primakoff 1945) but in its simple form it was developed independently of this result.

* Security reasons still prevent the publication of actual numerical data, but the principles involved are of general interest in acoustical work and it has been thought worth while to prepare the theoretical work for publication.

The reciprocity principle presents the concept that two transceivers placed in a sound-transmitting medium are equivalent to an electric quadripole, and that the transfer impedance is therefore independent of the input direction. In other words, the ratio of open circuit volts on the receiver to driving current in the projector is unique. A concomitant property of the system is that the ratio of the receiving sensitivity of a transceiver, measured in open circuit volts per unit pressure, to its projection sensitivity, measured in pressure units at a given distance per unit driving current, is also unique. This ratio has been called the "reciprocity parameter" and is denoted by J .

The basic principle of the method described in this paper circumvents the reciprocity principle, in effect, by making the analysis rest on voltage measurements together with self impedance measurements on the transceivers. Nevertheless, the full analysis of the method, as we shall see, cannot be carried out without invoking the reciprocity principle.

§ 2. THE BASIC EXPERIMENT

Let us assume that we have a perfect transceiver, 'T', for which axial pressure is required. We shall require an auxiliary hydrophone, H, and a projector, P, for producing a sound field. The hydrophone may be of any form whatever, provided that it is linear, and it may be supposed to be followed by an amplifier and an output meter. If the output meter records a voltage u when the hydrophone is placed in a field of which the undisturbed pressure is p , then we may write

$$p = hu, \quad \dots\dots(2.1)$$

where h is the calibration constant of the hydrophone.

The transceiver, 'T', may be represented as in figure 1, in which I is the motional impedance of the device, e is the voltage produced in it by the sound pressure when it is acting as a receiver, and C is a capacitative load. The first part of the basic experiment is to establish a sound field using the projector P, and to determine the voltage developed across C when 'T' is placed in this field. If p_w be the free field pressure, we have *

$$e = \frac{2A}{k} p_w, \quad \dots\dots(2.2)$$

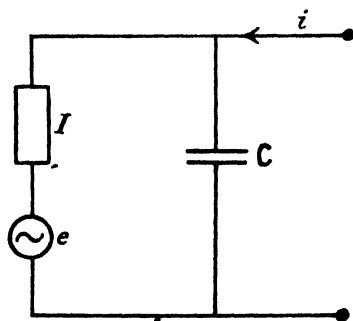


Figure 1.

where A is the effective area of the face of 'T' and k is a constant involving the piezo-electric constant of the crystal and the dimensions of the assembly.

The output voltage, v , developed across C is

$$v = \frac{e}{1 + KI} = \frac{2A}{k} \cdot p_w \frac{1}{1 + KI}, \quad \dots\dots(2.3)$$

where K is the conductance of C .

Since the measured value of v gives no indication of phase, the modulus of the impedance factor $1 + KI$ must be used.

* See (3.1) and (3.2), *infra*.

If, now, T is replaced by H, the output meter of H will record a voltage u_ω given by

$$p_\omega = hu_\omega. \quad \dots\dots(2.4)$$

Let T now be used as a projector, a driving voltage V being applied to its terminals. The power, W , radiated acoustically will be

$$W = \frac{v^2}{|I|^2} S, \quad \dots\dots(2.5)$$

where S is the radiation resistance. If the dimensions of the radiating face of T are not small compared with the wavelength, λ , the axial pressure, p , at a distance d ($\geq \lambda$) is given by*

$$p = \frac{k}{\lambda d} \sqrt{S W} = \frac{k}{\lambda d} \cdot \frac{VS}{|I|}. \quad \dots\dots(2.6)$$

If this pressure be recorded on H, an output voltage u will be produced, given by

$$p = hu. \quad \dots\dots(2.7)$$

From (2.3) and (2.4) we have

$$\frac{v}{u_\omega} = \frac{2Ah}{|1 + KI| k} = m_1. \quad \dots\dots(2.8)$$

From (2.6) and (2.7) we have

$$\frac{u}{V} = \frac{kS}{\lambda d |I| h} = m_2. \quad \dots\dots(2.9)$$

From (2.8) and (2.9) we find

$$m_1 m_2 = \frac{2AS}{\lambda d |I| |1 + KI|}, \quad \dots\dots(2.10)$$

$$\frac{m_1}{m_2} = \frac{2A\lambda d}{k^2 S} \frac{1}{h} \frac{1}{1 + KI}. \quad \dots\dots(2.11)$$

For a transceiver with a radiating face of dimensions not small compared with λ , we have (Vigouroux 1947)

$$k^2 S/A = \rho c, \quad \dots\dots(2.12)$$

where ρ is the density of the medium in which T is placed and c is the velocity of sound in the medium.

If we can ensure that $KI \gg 1$, (2.10) and (2.11) become

$$m_1 m_2 = \frac{2AS}{|K| \lambda d |I|^2} = \frac{2k^2}{|K| \lambda d \rho c} \cdot \frac{S^2}{|I|^2}, \quad \dots\dots(2.13)$$

$$m_1/m_2 = \frac{2\lambda d}{\rho c |K|} h^2. \quad \dots\dots(2.14)$$

The former relation depends only on the properties of T but not of H, and the latter on the properties of H but not of T. In particular, (2.14) allows us to determine h , the hydrophone constant, directly in terms of voltage ratios, impedance and frequency. From (2.6) and (2.13) we have, further,

$$\begin{aligned} \frac{p}{V} &= \frac{k}{\lambda d} \cdot \frac{S}{|I|} = \frac{k}{\lambda d} \sqrt{\frac{m_1 m_2 \lambda d \rho c |K|}{2k^2}}, \\ &= \sqrt{\frac{m_1 m_2 f \rho |K|}{2d}}, \end{aligned} \quad \dots\dots(2.15)$$

where $f = c/\lambda$ is the frequency of the sound. Thus the axial pressure is determined in terms of directly measurable quantities.

* See (3.1) and (3.2), *infra*.

and let the open circuit voltage v produced by a free field pressure ϖ_2 be given by

$$v = v\varpi_2. \quad \text{.....(3.12)}$$

Then the reciprocity theorem states that

$$\frac{v}{\mu} = J = \frac{2d\lambda}{\rho c}. \quad \text{.....(3.13)}$$

In figure 1 the actual current entering I is

$$i_1 = \frac{1}{1 + KI} i,$$

and the power radiated acoustically is

$$\begin{aligned} W' = S |i_1|^2 &= \frac{S}{|1 + KI|^2} i^2, \\ &= \rho c A_1 \dot{\xi}^2. \end{aligned}$$

The axial pressure for an ideal radiator is

$$\begin{aligned} p_0 &= \frac{\rho c A}{\lambda d} \dot{\xi}, \\ &= \frac{\rho f i}{d} \cdot \frac{1}{|1 + KI|} \sqrt{\frac{SA^2}{\rho c A_1}}. \end{aligned} \quad \text{.....(3.14)}$$

The actual axial pressure p_1 is

$$p_1 = \beta p_0 = \frac{\beta \rho f i}{d} \cdot \frac{1}{|1 + KI|} \sqrt{\frac{SA^2}{\rho c A_1}}. \quad \text{.....(3.15)}$$

Thus

$$\mu = \frac{\beta \rho f}{d} \cdot \frac{1}{|1 + KI|} \sqrt{\frac{SA^2}{\rho c A_1}}, \quad \text{.....(3.16)}$$

and

$$v = \frac{2d\lambda}{\rho c} \cdot \mu = 2\beta \frac{1}{|1 + KI|} \sqrt{\frac{SA^2}{\rho c A_1}}. \quad \text{.....(3.17)}$$

Thus, in a free field p ,

$$\begin{aligned} v &= v p, \\ &= 2\beta p \frac{1}{|1 + KI|} \sqrt{\frac{SA^2}{\rho c A_1}}. \end{aligned} \quad \text{.....(3.18)}$$

If e be the generated voltage

$$v = \frac{1}{|1 + KI|} e,$$

so that

$$e = 2\beta p \sqrt{\frac{SA^2}{\rho c A_1}}. \quad \text{.....(3.19)}$$

For an ideal piston the generated voltage is given by

$$e_0 = 2p \sqrt{\frac{SA^2}{\rho c A_1}} = e/\beta, \quad \text{.....(3.20)}$$

so that the generated voltage for an imperfect receiver is equal to its directivity coefficient multiplied by the generated voltage for an ideal receiver of the same materials and dimensions.

§ 4. THE TRIMETRIC SCHEME: HYDROPHONE CALIBRATION

We can now develop more fully the analysis given in § 2 for an ideal piston. We have, in fact, to rewrite the equations of that section in the light of the discussion given in § 3.

We have instead of (2.3)

$$v = \frac{2\beta A}{k} p_w \frac{1}{|1 + KI|}, \quad \dots\dots(4.1)$$

and instead of (2.6)

$$p = \frac{\beta k V S}{\lambda d} \frac{A}{|I| A_1}. \quad \dots\dots(4.2)$$

Combining these with (2.4) and (2.7) we have

$$m_1 = \frac{v}{u_w} = \frac{2\beta A h}{|1 + KI| k}, \quad \dots\dots(4.3)$$

$$m_2 = \frac{u}{V} = \frac{\beta k S}{\lambda d} \frac{A}{|I| h A_1}, \quad \dots\dots(4.4)$$

whence

$$m_1 m_2 = \frac{2\beta^2 A^2 S}{\lambda d A_1 |I| |1 + KI|} \quad \dots\dots(4.5)$$

$$\frac{m_1}{m_2} = \frac{2A_1 \lambda d}{k^2 S} \frac{1}{1 + KI} h^2. \quad \dots\dots(4.6)$$

Instead of (2.12) we must use the more precise form

$$k^2 S / A_1 = \rho c, \quad \dots\dots(4.7)$$

so that

$$h^2 = \frac{m_1}{m_2} \frac{1 + KI}{I} \frac{\rho f}{2d}, \quad \dots\dots(4.8)$$

which we may write as

$$h^2 = \frac{m_1}{m_2} \frac{|K| \rho f}{2d} \frac{1}{1 + \frac{1}{KI}} \quad \dots\dots(4.9)$$

Thus the calibration of the hydrophone does not depend on the directivity coefficient of T nor on the relation between its effective and geometrical areas, nor does it depend on measurements being made at or near the resonant frequency of T. If the load on T is a capacitance C, as in figure 1, having a power factor δ , then $|K|$ can be written as ωC to an accuracy of 0.5 δ^2 , and since δ is never likely to exceed 5%, we can in general ignore this correction.

The correction $1/KI$ will be small even if we work at the resonant frequency of T, for then I will approximate to the radiation resistance. This suggests that if the correction be not very large, use can be made of the ideal equation (2.16), which gives,

$$|K| S = \frac{2A}{\lambda d m_1 m_2},$$

so that, bearing in mind that we are dealing with r.m.s. quantities throughout, we have

$$h^2 = \frac{m_1}{m_2} \frac{\pi C \rho f^2}{d} \left(1 + \frac{m_1^2 m_2^2 \lambda^2 d^2}{8 A^2} \right). \quad \dots\dots(4.10)$$

We shall see, however, in general that $[1/KI]^2$ is in fact negligible compared with unity.

§ 5. THE TRIMETRIC SCHEME: TRANSCIEVER CALIBRATION

To determine the axial pressure of T we combine (4.2), (4.5) and (4.7), and obtain

$$p = V \sqrt{\frac{m_1 m_2 \rho f |K|}{2d}} \left(1 + \frac{1}{KI} \right) \quad (5.1)$$

This again does not depend on β , since the product $m_1 m_2$ is experimentally determined, nor on the ratio of effective and geometric areas.

Disregarding the corrections for $1/KI$ and δ , in the light of what has already been discussed, we have

$$p = I \sqrt{\frac{\pi m_1 m_2 \rho C}{d}} \quad \dots\dots (5.2)$$

Both this result and (4.10) are independent of any resonance conditions, and therefore we can measure absolutely the axial pressure at any frequency and so plot the true acoustic projector response curve.

We saw in § 2 that the possibility of determining radiation resistance arises from the present series of measurements (equation (2.16)). Actually, it appears from the more complete expression given in (4.5) that the particular property of the transceiver which does exist and which is directly determinable from this trimetric series of measurements is the quantity $\beta^2 A^2 S / A_1 |I|^2$.

In order to find what significance can be attached to this quantity, suppose that measurements are made at the resonant frequency of T—that is, at the frequency at which p/V , defined by (5.2), is a maximum. At this frequency $|I|$ becomes a pure resistance, the value of which may be denoted by S_1 . This will in general be greater than the radiation resistance S , because it will include internal losses arising from mechanical and other causes. Any electrical method which determines total power loss in the transceiver can be used to determine S_1 for, clearly, at a known r.m.s. driving voltage V the total power consumed will be V^2/S_1 . At resonance, therefore, we can determine a quantity R given by

$$R = \frac{A_1 S_1^2}{A^2 \beta^2 S} = \frac{A}{m_1 m_2 \lambda d \pi f C} \quad \dots\dots (5.3)$$

If the whole power consumed were radiated acoustically by an ideal piston of effective area A , identical with its geometrical area, the axial pressure p_0 at distance d would be given by

$$p_0^2 = \frac{\rho C A}{\lambda^2 d^2} \cdot \frac{V^2}{S_1} \quad \dots\dots (5.4)$$

The actual axial pressure, p , is, however, given by (5.2)

$$\begin{aligned} p^2 &= V^2 f^2 \frac{\pi m_1 m_2 \rho C}{d}, \\ &= \frac{V^2 f \rho A}{\lambda d^2 R} = \frac{V^2 \rho C A}{\lambda^2 d^2 R}. \end{aligned} \quad \dots\dots (5.5)$$

Thus

$$\frac{p^2}{p_0^2} = \frac{S_1}{R} \quad \dots\dots (5.6)$$

Thus with a knowledge of R and S_1 , we can determine the ratio of the actual axial intensity to that which would be given by an ideal piston radiating the whole

of the power as sound. This ratio may be called the "projection efficiency", and we can denote it by η_p^* . Thus

$$\eta_p = \frac{S_1}{R} = \frac{\pi d C c S_1 m_1 m_2}{A} \quad \dots\dots(5.7)$$

It is obvious that the efficiency with which T converts electrical power into sound power is given by the ratio S/S_1 . This may be called the "conversion efficiency", η_c , and we have

$$\eta_c = \frac{S}{S_1} = \eta_p \frac{A_1}{\beta^2 A} \quad \dots\dots(5.8)$$

Thus the actual conversion efficiency, or the actual sound energy radiated, cannot be determined from this scheme of measurements unless β and A_1/A are known. There is no accurate method of determining β : for axially symmetrical systems an estimate can be made from directional curves, but the method is unsatisfactory. Again, theoretical estimates of A_1/A rest on the assumption that β is unity, so that it appears that η_c is a somewhat inaccessible quantity.

This, however, need not concern us unduly, for the projection efficiency is an equally useful practical concept for asdic applications. While it is obvious that $\eta_c \geq 1$, yet it is not immediately clear that $\eta_p \geq 1$. This has led to the enunciation of the following proposition:—

"A plane radiator consuming, and radiating the whole of a given amount of power produces a maximum axial intensity at a given distance when it is vibrating as an ideal piston."

No formal proof has been evolved, but there is a strong supposition that the proposition is true. If it is true, then of course the upper limit of η_p would be unity.

§ 6. CONCLUSION

It has been shown that the scheme of measurements discussed in this paper can be made to yield the following results:—

- (i) The calibration constant for a hydrophone in "absolute" units.
- (ii) The axial pressure, at any frequency, of a transceiver, in "absolute" units.

These results depend only on directly measurable quantities such as voltage, capacitance, frequency and range. If measurements are made at the resonant frequency of a transceiver, then there is determined also:

- (iii) The projection efficiency.

It is hoped that it may be possible later to publish practical details of the technique developed from this paper. The following results indicate the order

* The choice of an ideal piston without edge-correction as a norm is arbitrary, and it is arguable that it would be better to take a simple piston with appropriate edge-correction as a reference standard. If this is done, we find that $p^2/p_0^2 = (S_1/R)(A_1/A)$, which is in general smaller than the quantity η_p as defined in (5.7). A third possibility is to take a small spherical radiator as a standard, but then the ratio of p^2/p_0^2 would depend on A/λ^2 , and no convenient physical interpretation can be placed on the result. The present suggestion for the definition of η_p is tentative, and may require revision as experience is acquired.

of magnitude of the quantities involved in the calibration of a small tourmaline hydrophone:—

$$\begin{aligned} m_1 &= 9.33 \cdot 10^1, & f &= 3.06 \cdot 10^4 \text{ c/s.}, \\ m_2 &= 1.08 \cdot 10^{-4}, & \rho &= 1.02 \text{ gr./cc.}, \\ C &= 4.01 \times 10^{-9} \text{ farad}, & S_1 &= 3.31 \cdot 10^4 \text{ ohms}, \\ d &= 6.10 \times 10^2 \text{ cm.} \end{aligned}$$

We note first that $1/KI = -j/2\pi \times 3.06 \cdot 4.01 \cdot 13.31 \cdot 10^{-1} = -0.04j$ so that $|1 + 1/KI|$ differs from unity by less than 1% . To this accuracy then we have, from (4.10)

$$h^2 = \frac{m_1}{m_2} \cdot \frac{\pi C \rho f^2}{d},$$

or

$$h = 5.70f \sqrt{\frac{m_1}{m_2} \cdot \frac{C}{d}} \text{ dynes/cm}^2 \text{ per volt,}$$

where C is in $\mu\mu\text{f}$., f is in kc/sec. , d is in cm. and $\pi\rho$ is embodied in the constant. Thus

$$\begin{aligned} h &= 5.70 \cdot 30.6 \sqrt{\left(\frac{9.33 \cdot 4.10}{1.08 \cdot 6.10} \cdot 10^6\right)}, \\ &= 4.20 \cdot 10^5 \text{ dynes/cm}^2 \text{ per volt.} \end{aligned}$$

This gives the hydrophone constant as already defined. Alternatively, the sensitivity can be expressed as $1/h$, i.e. as $2.38 \mu\text{v. per dyne/cm}^2$

It can be further shown that the quantity R (equation (5.3)) is, for the present cases, $6.15 \cdot 10^4$ ohms, and $S\beta^2 A/A_1 = S_1^2/R \cdot 18000$ ohms. Approximately, $A \approx A_1$, so that $S\beta^2 \approx 18000$ ohms, which means that the actual radiation resistance is of this order, although nothing is known of β except what might be inferred roughly from directional curves. Nevertheless, the example does serve to bring out the significance of the three "resistances" associated with a transceiver. The true radiation resistance, S , is that arising from the damping due to the medium (in our case, water), the dissipation resistance S_1 is the equivalent shunt resistance responsible for the absorption of electrical power, while the third quantity, which we have denoted by R , can be regarded as the radiation resistance of the equivalent perfect piston. The magnitudes will be in the order $R \sim S_1 \gg S$.

ACKNOWLEDGMENTS

The work described in this paper was carried out while the author was serving in the Royal Naval Scientific Service under Sir Charles Wright, to whom thanks are due for permission to publish it. It is a pleasure for the writer to record his indebtedness to Mr. J. Anderson, O.B.E., Chief Scientist at H.M. Underwater Detection Establishment and to Dr. Paul Vigoureux, Superintending Scientist at H.M. Torpedo Experimental Establishment, for the benefit of many stimulating discussions.

REFERENCES

- FOLDY, L., and PRIMAKOFF, H., 1945, *J. Acoust. Soc. Amer.*, **17**, 109.
 McLACHLAN, N. W., 1934, *Loud Speakers*, p. 96 (Oxford: Clarendon Press).
 McLACHLAN, N. W., 1941, *Bessel Functions for Engineers*, p. 98 (Oxford: University Press).
 MORSE, P. M., 1936, *Vibration and Sound*, p. 256 (New York: McGraw-Hill).
 PRIMAKOFF, H., and FOLDY, L., 1947, *Ibid.*, **19**, 50.
 RAYLEIGH, Lord, 1929, *Theory of Sound*, vol. 2, para. 302; p. 164 (London: Macmillan).
 VIGOUREUX, P., 1947, *Proc. Phys. Soc.*, **59**, 19.

A Method of Computing a Vertical Section of the Combined Polar Diagram of a Radio Aerial, a Flat Earth and a Vertical Screen

BY N. CORCORAN * AND J. M. HOUGH †

Malvern

* Now Senior Assistant, Stockport College for Further Education

† Now at University College, Hull

M.S. received 7 March 1947

ABSTRACT. The method described is based on the Sommerfeld formula for diffraction at an edge, combined with the effect of reflection at the earth. A table is given which reduces the amount of numerical work involved.

§ 1. INTRODUCTION

IN the winter of 1944-45 it was found necessary to use narrow-beam radar equipments at wavelengths of the order of 10 centimetres for locating targets at small angles of elevation. As the edge of the beam struck the ground near the equipment, there was a considerable amount of "ground-clutter" which greatly increased the background of noise as seen on the display, thus swamping the echo from the target. With a view to reducing this clutter, calculations and experiments were carried out on the use of opaque vertical screens placed in front of the equipment, providing an artificial horizon slightly above the true horizon. The experiments were carried out by Army personnel in the campaign in Holland, and by Hey, Parsons and Jackson of Army Operational Research Group in this country.

By a well-known theorem, the polar diagram of an aerial system is the same for transmitting as for receiving; it is more convenient to consider the equipment as a receiver. The treatment adopted is to sum with respect to phase and amplitude all the rays which reach the aerial, deducing the relative phases from the path differences, and the relative amplitudes from the Sommerfeld expression for the amplitude diffracted through a given angle.

§ 2. THEORY

The arrangement of screen SS' and receiver or transmitter T considered is shown in figure 1. The calculations are simplified by considering T to be a receiver and by considering plane waves coming from the right at an angle θ . In all there are ten rays to consider, which can be put into four groups. Group 1 consists of rays which have suffered no reflection; Group 2 consists of rays which have been reflected before reaching the screen; Group 3 consists of rays which have been reflected after reaching the screen; and Group 4 consists of rays which have been reflected both before and after reaching the screen.

The component rays in each group are:

- Group 1. Direct ray TN_1 if $\theta > A$;
Diffracted rays TSN_2 and $TS'N_3$.
- Group 2. Reflected ray TQ_1N_4 if $D > \theta > B$;
Diffracted rays TSQ_3N_6 and $TS'Q_2N_5$.
- Group 3. Reflected ray TQ_1N_4 if $\theta > C$;
Diffracted rays TQ_4SN_2 and $TQ_5S'N_3$.
- Group 4. Diffracted rays $TQ_4SQ_3N_6$ and $TQ_5S'Q_2N_5$.

All rays in a given group arrive at the equipment in phase except for the phase changes on diffraction.* If the resulting complex amplitude is A_1 for Group 1, A_2 for Group 2, etc., and the reflection coefficient is R , we have:

$$\text{Resultant field} = A_1 + A_4 R^2 + (A_2 + A_3) R e^{-i4\pi h_T \sin \theta / \lambda}.$$

The reflection coefficient R is in general a complex quantity which varies with the angle of incidence. For horizontal polarization R does not differ greatly

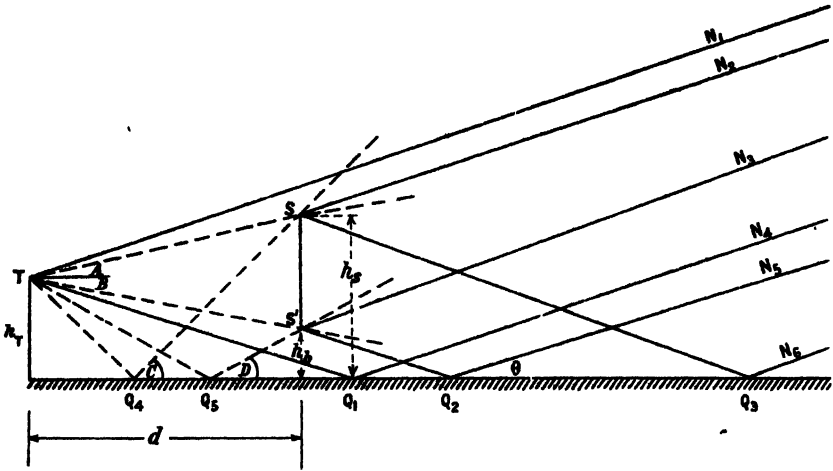


Figure 1.

from -1 , and for vertical polarization and small values of θ this is also approximately true. In this problem the angles are always small, and so R is taken to be -1 .

$$\therefore \text{Resultant field is } A_1 + A_4 - (A_2 + A_3) e^{-i4\pi h_T \sin \theta / \lambda}.$$

If the free-space polar diagram of the equipment over the range of angles concerned is appreciably different from that of an isotropic source, the rays should be multiplied by a correcting factor $F(\theta)$ for the direct ray, $F(-\theta)$ for the reflected ray, and $F(A)$, $F(-B)$, $F(-C)$ and $F(-D)$ for the diffracted rays, where $F(\theta)$ is the amplitude of the voltage polar diagram in the direction θ .

* The phase is calculated on the basis of the path-length of the geometrical-optics ray (real or virtual) for the given group, and not that of the ray which travels via the diffracting edge; this is not of course the actual phase of the diffracted ray, or of the resultant of the group, because of the phase change due to diffraction.

To calculate the diffracted rays, use is made of Sommerfeld's solution of the half-plane. In his solution there are two terms, one of which corresponds to simple diffraction and the other to diffraction and reflection. This second term is always small and is ignored. Figure 2 is a simplified part of figure 1, and the diffracted ray TSN_2 is given by Sommerfeld as

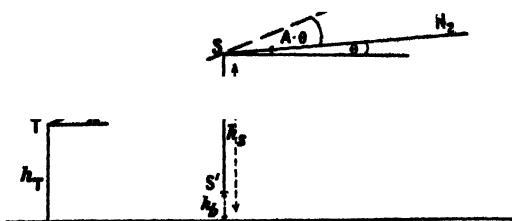


Figure 2.

$$\left[\frac{1}{2} + \frac{(1+i)}{2} (C - iS) \right],$$

where $(C - iS)$ is the Fresnel integral function for

$$u = 2 \left(\frac{2d}{\lambda} \right)^{\frac{1}{2}} \sin (A - \theta)$$

The expression is taken positive when the diffracted ray is in the shadow region and negative when in the illuminated region, i.e. the sign of the function is the same as the sign of u .

Similar expressions are found for the other rays. The values of u for the rays are

Ray	$u / 2 \left(\frac{2d}{\lambda} \right)^{1/2}$
Group 1 $\left\{ \begin{array}{l} TSN_2 \\ TS'N_3 \end{array} \right.$	$\sin \frac{A - \theta}{2}$ $\sin \frac{B + \theta}{2}$
Group 2 $\left\{ \begin{array}{l} TSQ_3N_6 \\ TS'Q_2N_5 \end{array} \right.$	$\sin \frac{A + \theta}{2}$ $\sin \frac{B - \theta}{2}$
Group 3 $\left\{ \begin{array}{l} TQ_4SN_2 \\ TQ_5S'N_3 \end{array} \right.$	$\sin \frac{C - \theta}{2}$ $\sin \frac{D - \theta}{2}$
Group 4 $\left\{ \begin{array}{l} TQ_4SQ_3N_6 \\ TQ_5S'Q_2N_5 \end{array} \right.$	$\sin \frac{C + \theta}{2}$ $\sin \frac{D + \theta}{2}$

§ 3. METHOD OF COMPUTATION

For simplicity of computation the Sommerfeld formula is split into real and imaginary parts:

$$\frac{(1+C)}{2} (C-iS) = \frac{1}{2} [(1+C+S) + i(C-iS)] \\ = M + iN \text{ say.}$$

C and S can be found in tables as functions of u , and table 1 gives M and N as functions of u . As the angles are small, we have approximately:

$$u = \frac{\pi}{180} \sqrt{\frac{2d}{\lambda}} F,$$

where F is the angle of diffraction and is given in degrees, or

$$u = 0.0247 \sqrt{\frac{d}{\lambda}} F;$$

$$\text{also } A = \tan^{-1} \frac{h_s - h_T}{d}, \quad B = \tan^{-1} \frac{h_T - h_b}{d}, \\ C = \tan^{-1} \frac{h_s + h_T}{d}, \quad D = \tan^{-1} \frac{h_T + h_b}{d}.$$

Thus the method of computing consists of choosing a value of θ , finding the corresponding values of u , and then adding real and imaginary parts of the five rays in Group 1 and Group 4 and also the five rays in Groups 2 and 3. If θ is chosen in such a way that

$$\frac{4\pi h_T \sin \theta}{\lambda} = \frac{n\pi}{2},$$

then the phase difference between the two sets of rays is a multiple of 90° . This simplifies the addition, as it only involves multiplying the resultant real and imaginary parts of Groups 2 and 3 by ± 1 or $\pm i$. The resultant intensity in the direction θ is then the sum of the squares of the real and imaginary parts. This process is repeated for different values of θ , thus giving the polar diagram. In practice some of the above rays give a negligible contribution to the total amplitude. An example is given in the Appendix.

§ 4. COMPARISON WITH EXPERIMENT

Figure 3 gives a set of readings taken for a GL3 with a screen 9 ft. from top to bottom, placed 35 yards from the set, the angle of elevation of the set being 2° and the top of the screen 15 ft. above ground. For the purpose of the calculations λ was taken as 10 cm. and the height of the transmitter as 12 ft. above ground. The calculated curve is shown in figure 14. It will be seen that the calculated points agree reasonably well with the observed curve, and it therefore appears that the approximate method outlined above gives a sufficiently good approximation for practical purposes in the field. The experimental curve was obtained by Mr. J. T. G. Milne, of the Army Operational Research Group, while temporarily attached to 21 Army Group.

It is important to note that the observations were made in Holland, where the ground is exceptionally flat over large areas, and that such agreement may not be found on more undulating country.

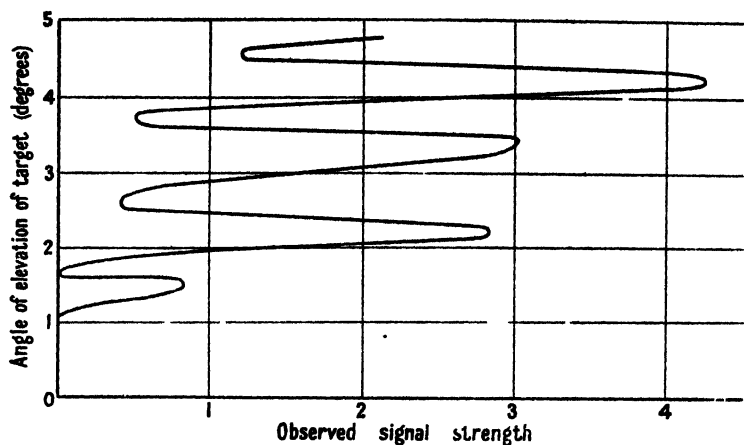


Figure 3. Signal strength of target echo above a wire screen (observed).

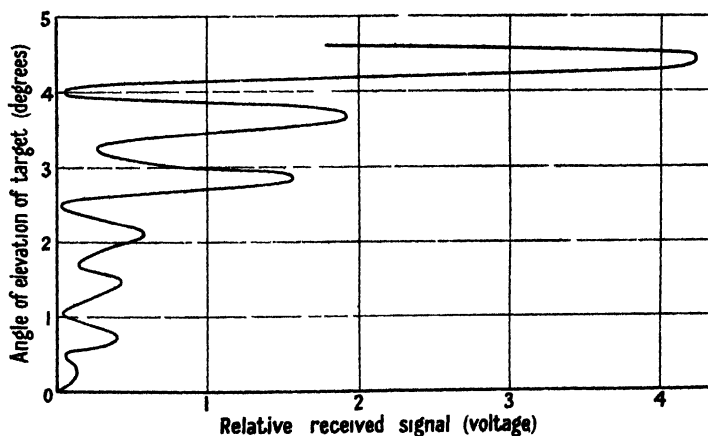


Figure 4. Signal strength of target echo above a wire screen (theoretical).

§ 5. ACKNOWLEDGMENTS

The above work was carried out while the authors were on the staff of the Radar Research and Development Establishment of the Ministry of Supply, and the authors wish to thank the Chief Scientist, Ministry of Supply, for permission to publish this paper. They also wish to thank Mr. J. M. C. Scott for his help and advice.

Table 1

<i>u</i>	<i>M</i>	<i>N</i>	<i>u</i>	<i>M</i>	<i>N</i>
0.000	0.500	0.000	1.300	--0.163	--0.023
0.025	0.488	0.012	1.325	- 0.156	-0.039
0.050	0.475	0.025	1.350	- 0.148	0.055
0.075	0.462	0.037	1.375	- 0.139	- 0.070
0.100	0.450	0.050	1.400	- 0.128	-0.083
0.125	0.437	0.062	1.425	-0.114	-0.095
0.150	0.424	0.074	1.450	-0.100	-0.106
0.175	0.411	0.086	1.475	-0.086	-0.116
0.200	0.399	0.098	1.500	-0.071	-0.124
0.225	0.385	0.109	1.525	-0.054	-0.131
0.250	0.371	0.120	1.550	-0.037	-0.134
0.275	0.357	0.131	1.575	-0.020	-0.136
0.300	0.343	0.142	1.600	- 0.003	- 0.137
0.325	0.329	0.152	1.625	0.014	-0.136
0.350	0.314	0.162	1.650	0.031	-0.133
0.375	0.299	0.172	1.675	0.047	-0.125
0.400	0.284	0.181	1.700	0.061	- 0.113
0.425	0.269	0.190	1.725	0.074	-0.099
0.450	0.253	0.198	1.750	0.086	- 0.086
0.475	0.237	0.205	1.775	0.097	-0.072
0.500	0.221	0.212	1.800	0.105	-0.057
0.525	0.205	0.219	1.825	0.111	-0.041
0.550	0.188	0.225	1.850	0.115	- 0.024
0.575	0.171	0.230	1.875	0.117	-0.007
0.600	0.154	0.236	1.900	0.116	0.011
0.625	0.137	0.240	1.925	0.114	0.026
0.650	0.119	0.242	1.950	0.109	0.042
0.675	0.102	0.243	1.975	0.100	0.057
0.700	0.084	0.244	2.000	0.084	0.072
0.725	0.067	0.244	2.025	0.069	0.084
0.750	0.049	0.243	2.050	0.053	0.093
0.775	0.032	0.240	2.075	0.038	0.100
0.800	0.014	0.236	2.100	0.022	0.104
0.825	-0.003	0.231	2.125	0.005	0.104
0.850	-0.019	0.225	2.150	-0.012	0.102
0.875	-0.037	0.219	2.175	-0.029	0.100
0.900	-0.052	0.211	2.200	-0.048	0.090
0.925	-0.067	0.202	2.225	-0.057	0.077
0.950	-0.081	0.192	2.250	-0.068	0.064
0.975	-0.095	0.181	2.275	-0.079	0.050
1.000	-0.109	0.169	2.300	-0.092	0.037
1.025	-0.121	0.157	2.325	-0.094	0.020
1.050	-0.132	0.143	2.350	-0.095	0.002
1.075	-0.142	0.128	2.375	-0.093	-0.015
1.100	-0.151	0.113	2.400	-0.087	-0.032
1.125	-0.156	0.097	2.425	-0.075	-0.045
1.150	-0.161	0.080	2.450	-0.063	-0.057
1.175	-0.165	0.063	2.475	-0.051	-0.069
1.200	-0.168	0.046	2.500	-0.038	-0.081
1.225	-0.170	0.029	2.525	-0.021	-0.086
1.250	-0.169	0.011	2.550	-0.004	-0.086
1.275	-0.167	-0.006	2.575	0.013	-0.084

Table 1 (contd.)

<i>u</i>	<i>M</i>	<i>N</i>	<i>u</i>	<i>M</i>	<i>N</i>
2.600	0.030	-0.081	3.900	0.051	-0.027
2.625	0.042	-0.068	3.925	0.056	-0.010
2.650	0.054	-0.055	3.950	0.058	0.006
2.675	0.066	-0.043	3.975	0.058	0.023
2.700	0.075	-0.030	4.000	0.041	0.039
2.725	0.079	-0.013	4.025	0.023	0.050
2.750	0.080	-0.004	4.050	0.005	0.054
2.775	0.078	0.021	4.075	-0.012	0.054
2.800	0.071	0.038	4.100	-0.030	0.049
2.825	0.056	0.056	4.125	-0.044	0.034
2.850	0.042	0.066	4.150	-0.054	0.019
2.875	0.028	0.073	4.175	-0.054	0.004
2.900	0.013	0.076	4.200	-0.053	0.011
2.925	0.003	0.076	4.225	-0.048	-0.030
2.950	0.019	0.072	4.250	-0.038	-0.046
2.975	-0.035	0.063	4.275	-0.024	-0.052
3.000	-0.051	0.055	4.300	-0.002	-0.052
3.025	-0.061	0.039	4.325	0.020	0.051
3.050	-0.069	0.022	4.350	0.038	-0.040
3.075	-0.073	0.006	4.375	0.048	-0.026
3.100	-0.072	-0.010	4.400	0.050	-0.012
3.125	-0.068	-0.023	4.425	0.047	0.002
3.150	0.058	-0.037	4.450	0.040	0.017
3.175	-0.044	-0.050	4.475	0.030	0.031
3.200	-0.030	0.064	4.500	0.020	0.046
3.225	-0.013	0.067	4.525	0.005	0.049
3.250	0.004	-0.068	4.550	-0.011	0.048
3.275	0.021	-0.067	4.575	-0.028	0.036
3.300	0.038	-0.057	4.600	-0.042	0.025
3.325	0.054	-0.042	4.625	-0.047	0.010
3.350	0.062	-0.026	4.650	-0.047	-0.006
3.375	0.065	-0.011	4.675	-0.045	-0.022
3.400	0.066	0.004	4.700	-0.029	-0.038
3.425	0.066	0.018	4.725	0.013	-0.045
3.450	0.058	0.032	4.750	0.003	-0.047
3.475	0.042	0.045	4.775	0.019	-0.042
3.500	0.026	0.059	4.800	0.035	-0.032
3.525	0.010	0.062	4.825	0.045	-0.015
3.550	-0.007	0.062	4.850	0.046	0.001
3.575	-0.024	0.060	4.875	0.045	0.017
3.600	-0.040	0.048	4.900	0.032	0.033
3.625	-0.052	0.032	4.925	0.017	0.045
3.650	-0.060	0.016	4.950	0.001	0.045
3.675	-0.063	-0.001	4.975	-0.016	0.042
3.700	-0.057	-0.017	5.000	-0.032	0.032
3.725	-0.046	-0.032			
3.750	-0.033	-0.046			
3.775	-0.020	-0.058			
3.800	-0.007	-0.059			
3.825	0.008	-0.058			
3.850	0.022	-0.050			
3.875	0.037	-0.038			

APPENDIX

Computation of one point

To illustrate the method of computation, the intensity for one angle of elevation is calculated in detail below. The following values of the relevant lengths are used:—

Transmitter height	12 ft.
Distance from screen to transmitter	100 ft.
Height of bottom of screen	6 ft.
Height of top of screen	15 ft.
Wavelength	4 in.

The four fixed angles are, therefore,

$$A = \tan^{-1} \frac{15-12}{100} = 1^\circ.7,$$

$$B = \tan^{-1} \frac{12-6}{100} = 3^\circ.45,$$

$$C = \tan^{-1} \frac{15+12}{100} = 15^\circ.1,$$

$$D = \tan^{-1} \frac{12+6}{100} = 9^\circ.65,$$

and the parameter $n = 0.0247 \sqrt{\frac{d}{\lambda}} F = 0.428 F$,

F being the angle of diffraction measured in degrees. For convenience we choose our steps of θ such that

$$\frac{4\pi \cdot 12 \cdot \theta}{\frac{1}{3}} = \frac{n\pi}{2},$$

i.e., θ increases in steps of 0.2 . This interval is sufficiently small for the resulting points to give a good indication of the curve.

As an example, the calculation for $\theta = 2^\circ.8$ is given. For this angle there is only one geometrical-optics ray, TN, and so we have the following table.

Group	Ray	F	u	M (from table 1)	N (from table 1)
1	TN ₁	0	0	1.000	0
	TSN ₂	-1.1	0.0471	0.240	-0.204
	TS'N ₃	6.25	2.674	0.066	-0.043
2	TSQ ₃ N ₆	4.5	1.925	0.114	0.026
	TS'Q ₂ N ₅	0.65	0.278	0.355	0.132
3	TQ ₄ SN ₂	12.3	5.262*		
	TQ ₅ S'N ₃	6.85	2.931	-0.007	0.075
4	TQ ₄ SQ ₃ N ₆	17.9	7.658*		
	TQ ₅ S'Q ₂ N ₅	12.45	5.326*		
				1.288	-0.014

* The table of M and N is only given for $0 \leq u \leq 5$. For larger u the diffracted ray is negligible owing to the narrow polar curve of the equipment.

The geometrical path of Group-2 rays is 3.5λ greater than that of Group 1. As these rays undergo a phase reversal on reflection, all groups arrive in phase and the summation above is correct. Had we taken $\theta = 2^\circ.4$, rays in Groups 2 and 3 would arrive in opposite phase to those in Groups 1 and 4, and would, therefore, be multiplied by -1 before being added.

LETTERS TO THE EDITOR

An Isotopic Abundance Rule and its Bearing on the Origin of the Nuclei

The natural abundances of all stable isotopes of any element of higher atomic number than 43 are of similar magnitude, except when there exist natural isobars of lower atomic number ; those that have isobars of lower charge are comparatively rare.

The limit at $Z = 43$ is rather arbitrary ; indications of the same type of abundance pattern persist down to ${}_{34}\text{Se}$ but are not present among light nuclei. "Of similar magnitude" means, commonly, within a factor of 2 ; at the worst, within a factor of 11 (${}^{138}_{56}\text{Ba} : {}^{138}_{58}\text{Ba}$). "Comparatively rare" means, generally, between 10 times and 100 times rarer. There are only four actual exceptions, i.e. nuclei which, possessing lower charge isobars, are more abundant than one or more of their isotopes which possess none. These are ${}^{110}_{48}\text{Cd}$, ${}^{116}_{50}\text{Sn}$, ${}^{142}_{60}\text{Nd}$, ${}^{198}_{80}\text{Hg}$. Of those isotopes which do not possess lower charge isobars, the heaviest is occasionally rather less abundant than the rest ; in all such cases it possesses a higher charge isobar, which is less rare than usual (the four exceptions mentioned are such cases).

The isotopic abundance pattern of ${}_{54}\text{Xe}$ (table 1) is particularly interesting, since, in obeying the above rule, it breaks the usual rule that every odd isotope of even atomic number has at least one more abundant adjacent even isotope.

The only other exceptions to the latter rule, beyond boron, are ${}^{147}_{62}\text{Sm}$, ${}^{148}_{62}\text{Sm}$, ${}^{196}_{78}\text{Pt}$ and ${}^{235}_{92}\text{U}$.

Table 1 (From Mattauch, *Kemphysikalische Tabellen* (Berlin : Springer))

${}_{54}\text{Xe}$ Mass No.	124	126	128	129	130	131	132	134	136
Abundance %	0.094	0.088	1.91	26.23	4.06	21.18	26.98	10.55	8.95
Lower charge isobar	Yes	Yes	Yes	No	Yes	No	No	No	No

In conjunction with Goldschmidt's (1937) estimates of the cosmic abundances of the elements, the first rule may be generalized to the form :

The total cosmic nuclear abundance at each mass number (above 60) is of the same order of magnitude ; where there is more than one stable nucleus of a given mass, the predominantly abundant nucleus at this mass is that of lowest charge, and the others may be much rarer.

The only grave exception to the rule in this form is the relative rarity of mass number 185 (${}^{185}_{75}\text{Re}$, possessing no isobar), by a factor between 100 and 1000. It is suggested that new determinations of the abundance of rhenium, particularly in meteorites, are called for.

The rule in this latter form gives evidence in support of the hypothesis that in the origin of the nuclei, the heavier ones (as distinct from the lighter nuclei whose abundance is known to correspond approximately to a thermodynamic equilibrium (von Weizsäcker 1938, Chandrasekhar and Henrich 1942)) have been built up by a sequence of neutron capture processes. Since the nuclei then assume their final charge chiefly by β -decay, the predominance of the lowest charge isobars is a natural consequence. Detailed considerations of the neutron condensation process (to be published elsewhere) account also for the general form of the abundance curve for the heavier nuclei, and show that this process is the natural sequel to the thermodynamic equilibrium state indicated by the cosmic abundances of light nuclei.

F. C. FRANK.

H. H. WILLS Physical Laboratory,
University of Bristol.

REVIEWS OF BOOKS

Achievements in Optics, by A. BOUWERS. Pp. viii+125. (Brit. Emp. Distr.: Cleaver Hume Press, 1947.) 12s. net.

This book, one of a series of monographs dealing with progress of research in Holland, was prepared for the press before the liberation of the Low Countries, and deals almost exclusively with work done—often under conditions of secrecy—during the occupation.

The first chapter deals with "New Optical Systems". A clear and simple account is given of the advantages possessed by the concave spherical mirror, viz. small spherical aberration, complete absence of chromatic aberration, negative Petzval sum and absence of all orders of oblique aberrations when used with a stop at the centre of curvature. A series of systems is then described, comprising the Schmidt camera, the Maksutov telescope, and derivatives of these two. Very notable results seem to have been achieved. It is interesting to note that yet another claimant to priority for the use of weak negative meniscus lens for correcting the aberrations of a concave spherical mirror has appeared. This device has been generally attributed, so far, both in America and in this country, to Maksutov.

Chapter II deals with applications of mirror systems to microscopes, telescopes and cameras, and contains little that will be unfamiliar to the informed reader.

Chapter III deals with geometrical aberration theory. An interesting discussion of purely geometrical constructions for ray-paths and image points shows that even this somewhat dated approach can yet yield useful results. It is followed by a summary of the work of Korringa (Thesis, Delft, 1942) and Stephan, and Nijboer (Thesis, Groningen, 1942), all of which starts from the analytical standpoint introduced by Hamilton and developed in England, in particular by T. Smith. The chapter concludes with a description of a method of dealing with the light distributions in images in the presence of larger amounts of spherical aberration. As with so many of the methods of applied mathematics, one can but say that the grave suspicions provoked by the analytical procedure are allayed by the apparent success of the method.

Chapter IV, "Physical Optics", summarizes the elegant treatment of aberrational diffraction theory given by Nijboer (*ibid.*) dealing with cases in which aberrational asphericities are small; and finally, Zernike's remarkable analysis of the theory of the microscope is given, with particular reference to the phase-contrast method.

The English used is but rarely foreign to a native ear. The only serious ambiguity arises when a full-circular aperture is described as ring-shaped. This would easily be taken to mean an annular aperture, and in the context the two have to be carefully distinguished. Apart from this, and the use of "feeded" and "grinded" for the past participles of "to feed" and "to grind" respectively, a very fluent and natural style is maintained.

In sum, this book offers little really new information, but a wealth of scattered and useful material has been compressed into it. It can be wholeheartedly recommended to specialists and non-specialists alike. Finally, it is no mere courtesy to congratulate the Dutch workers who have contributed to optical research with such enviable enthusiasm and pronounced success under very unfavourable conditions.

H. H. H.

Microtecnic. International review for measuring and gauging technique, optics and precision mechanics. (Published in Lausanne, Switzerland.) Price 52s. 6d. per annum. British agents: Bailey Bros and Swinfen, London.

A bi-monthly journal with text in both English and French. A specimen issue includes articles on:—Documentation and International Decimal Classification; Some Applications of Optics to Metrology; The Hairspring; Equipment of a Laboratory for Precision Instruments; Beryllium and Beryllium Bronze; Rotating substandard Meters and their Uses.

A. C. S.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 60, PART 3

1 March 1948

No. 339

The Adiabatic Temperature Changes Accompanying the Magnetization of some Ferromagnetic Alloys in Low and Moderate Fields

BY L. F. BATES AND E. G. HARRISON

University College, Nottingham

MS. received 30 May 1947; read 8 November 1947

ABSTRACT. The method of Bates and Weston for measuring the small heat changes which occur as a ferromagnetic substance is taken through a single hysteresis cycle step by step has been extended to some seven ferromagnetic alloys. In every case a cooling is initially observed as the magnetization of the specimen is reduced from its maximum value. Indications are found that large changes in thermomagnetic properties are caused by small changes in composition, particularly when the latter approximates to that of invar. The results are discussed in terms of the modern domain concepts of ferromagnetic processes.

§ 1. INTRODUCTION

THE temperature changes which occur when a ferromagnetic substance is taken through an ordinary or so-called 'technical' hysteresis cycle were measured by a new method by Bates and Weston (1941) and described in a paper hereafter termed Paper I. In this paper a fairly comprehensive study was made of the behaviour of pure nickel in the annealed, hard-drawn and externally stressed conditions, and a few measurements on certain nickel-iron alloys such as invar, mu-metal and permalloy were described. Later, Bates and Healey, in Paper II (1942), studied the behaviour of Armco iron, and, more recently, Bates and Edmondson in Paper III (1947) have published results for cobalt.

References to earlier work by other investigators and a full description of the experimental arrangements and technique were given in Paper I and reviewed briefly in Papers II and III, so that only the barest details need be given here. The rod specimen was mounted vertically along the axis of a water-cooled solenoid supplied by a steady current which could be changed in steps. Some twenty separate copper-constantan thermocouples were arranged so that the 'hot' junction of each couple was in intimate contact with the rod, while the corresponding 'cold' junction was close to, but thermally insulated from, the rod. Each couple was joined in series with its own low-resistance primary winding, wound upon a mu-metal spiral core or ring. A low-resistance secondary coil of some 2000 turns was also wound upon the core and connected to a specially designed fluxmeter of high sensitivity. When the temperature of the rod was suddenly changed by ΔT , a current proportional to ΔT flowed for a short interval of time in each of the primary windings, so that the change of flux through the secondary

circuit caused a ballistic deflection of the fluxmeter proportional to ΔT . The instrument could usually be calibrated by producing known changes of temperature by suddenly applying a longitudinal stretching force to the rod.

§2 EXPERIMENTAL PROCEDURE

The magnetic hysteresis cycle was measured by the usual ballistic method, using a search-coil wound over the middle portion of the rod, all measurements being standardized by a calibrated Grassot fluxmeter and by a standard mutual inductance. The sources of error peculiar to the thermal measurements have been discussed in the three earlier papers, and, in the present work, the same steps were taken to avoid errors due to zero drift, lack of thermal insulation, eddy currents set up in the specimen and the effects of stray fields from the solenoid and specimen upon the mu-metal core and upon the thermocouple circuits as were used on the previous occasions.

The fluxmeter, with its electromagnetic field control, was overhauled, and its sensitivity reached that recorded by Bates and Weston, with a steady zero. In addition, we were fortunate in having at our disposal two Tinsley ballistic moving coil galvanometers of low resistance, long period and high sensitivity. One of these, with a coil resistance of 16.5 ohms, was found to be very suitable as a fluxmeter for our work, provided that its permanent magnet was suitably shunted. The makers gave 350 ohms as its critical damping resistance under their standard conditions. It was necessary to reduce this value considerably to be more in line with the resistance of the secondary of the mu-metal core, which was about 30 ohms. By the use of a magnetic shunt the sensitivity for our purposes was increased by some 80% and this figure could have been much increased by the provision of a new secondary with very many more turns.

The Tinsley instrument was very much less 'temperamental' than our own fluxmeter, and it had practically no zero drift. Its lens window was removed and a thin, carefully selected, mica disc was substituted. By very good fortune the mirror, which was of good quality, was slightly convex. 'This enabled the telescope to be focused on a virtual image considerably nearer to the object glass than would have been the case for the image in a plane glass mirror.' The magnified scale divisions were therefore seen with excellent sharpness and clarity, particularly when a narrow rectangular stop was placed before the mirror. The increased magnification allowed the scale to be placed further from the mirror at a distance limited by the dimensions of the research laboratory, namely 9 metres. In these circumstances the sensitivity was about 80% of that of the original fluxmeter as used by Bates and Weston, but the new instrument could, of course, give the same sensitivity if it were used in a longer laboratory. It follows that efforts will be made to provide larger secondaries and longer laboratories for future work. In the present work the new instrument was used in our experiments on 42% nickel-iron and mangonic.

Reference has already been made to the calibration of the temperature measuring system by the Joule loading method. This is excellent, provided that the specimen under examination possesses a reasonably large and accurately known coefficient of linear expansion. If this is not the case, we may assume that Warburg's law holds, and measure the deflection corresponding to the complete

traversing of one cycle which we may then equate to the temperature rise calculated from $\oint H dI$. But, in some alloys, the coefficient of linear expansion is very low, and the value of overall heating so small in comparison with the observed heat changes in certain parts of the cycle, that the overall heating may not be deduced with confidence. It was therefore desirable to have a check on the values obtained by both these methods, and the following was devised.

It was found that the fluxmeter showed no deflection when comparatively large alternating currents, of the order of milliamperes, were sent through the primary windings of the mu-metal core. Hence, while it was impracticable to heat the specimen by direct current, because of 'pickup' effects with the thermocouples, it could be heated by passing an alternating current through it for a known short time, when the fluxmeter showed a deflection corresponding to the rise in temperature. The resistance of the specimen per cm. length was found and the R.M.S. current and the time of flow measured.

For the latter measurement the circuit of figure 1 was used. The alternating current was obtained from a 50:1 step-down transformer connected to the mains

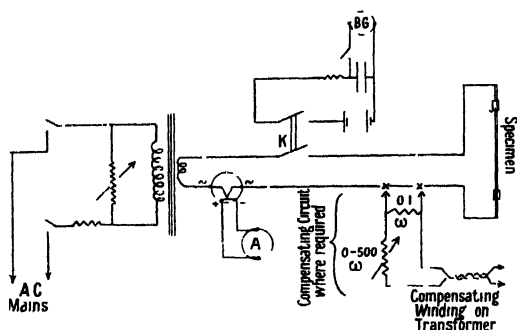


Figure 1. Circuit for A.C. calibration of the temperature-measuring system.

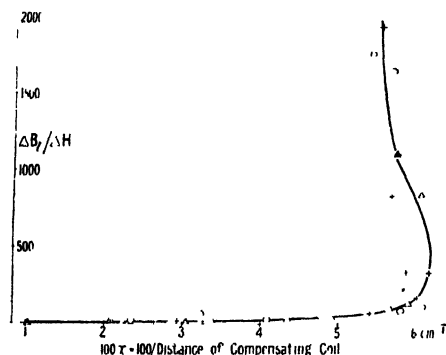


Figure 2. The behaviour of the induction effect for 36Ni-64Fe alloy.

via the resistance network shown. The current was measured by a calibrated thermo-ammeter and entered the specimen through the rigid double mercury key K, carefully arranged so that the heating circuit and the time-measuring circuit were made and broken simultaneously; the heating current was allowed to flow for about 1 second. The time-measuring circuit was of the familiar condenser discharge type. A compensating circuit was found necessary in the case of specimen W.5 to overcome rectification phenomena or pickup by the thermocouples which resulted in a sharp inductive deflection of the fluxmeter; with all other specimens this was not needed.

It is assumed, of course, that the passage of alternating current through the specimen does not produce hysteresis or eddy-current heating. This is the case when the circular magnetization it produces in the specimen is less than the longitudinal magnetization produced in the specimen by an alternating current, supplied to the solenoid, for which the fluxmeter just fails to record a heating of the specimen. In practice, a smaller current than that given by the foregoing

condition was used. Very satisfactory agreement between the three methods of calibration was obtained in all our work. We have, however, recently found that the alternating-current method is unreliable with specimens of pure iron, but it is easy to provide an explanation for the lack of agreement in these cases.

It remains now to mention a few points in which the experimental details differed from those of Papers I to III. The thermocouples were renewed and their individual resistances decreased by shortening the constantan strips; this brought the 'hot' and 'cold' junctions closer together and decreased zero drift. The hot junctions were now lightly bound to the rod with doubled waxed thread, and, since the binding completely covered the tip of each junction, temperature fluctuations and zero drift were thereby reduced. The rate of change of solenoid current with time was cut down by the insertion of a 2-henry choke in the solenoid circuit; this assisted in the reduction of eddy-current heating. The choke was, of course, shorted while the specimen was being demagnetized. Cooling water was not taken direct from the mains, but was siphoned from a reservoir through coils of lead tubing immersed in a second reservoir, so that it reached a constant temperature independent of mains fluctuations.

With some specimens, of large diameter, eddy-current heating was apparent. Its contribution to the total heating observed was estimated for each cycle from the mean thermal deflection, d , observed when the field was changed in a single step from zero to the maximum value. The sum of the deflections, Σd_1 , obtained by describing the same change in magnetization in a number of steps, was subtracted from d . We know the square of the change of induction $(\Delta B)^2$, corresponding to α , and the sum of the squares of the changes of induction, $\Sigma(\Delta B_1)^2$, corresponding to Σd_1 . Now, $[d - \Sigma d_1]$ must be proportional to $[(\Delta B)^2 - \Sigma(\Delta B_1)^2]$, so we can at once obtain the factor, p , say, which enables us to correct any individual reading, such as d_1 , for eddy-current heating. Thus, when we observe the deflection d_1 for a known change in induction ΔB_1 , then, in order to correct for eddy-current heating effects, we subtract from d_1 the quantity $p \cdot (\Delta B_1)^2$.

Bates and Weston used a compensating coil, denoted by C in figure 4 of Paper I, to try to equalize the fluxmeter deflections obtained on traversing a given step in magnetization in the two possible directions. It was found that one asymmetry arose from induction between the solenoid and mu-metal core, which could not be lessened by placing the latter in other positions with respect to the solenoid; fortunately, it is rarely serious. Another, and much more troublesome, asymmetry was termed the *induction effect*; it appears to arise from induction between the specimen and the thermocouple materials themselves.

We examined this matter in some detail, and in figure 2 are plotted the values of $\Delta B_s/\Delta H$ against the reciprocal, x , of the distance of the compensating coil from the solenoid when exact compensation was attained, the solenoid field change, ΔH , being kept constant. For, in these circumstances, x is a measure of the asymmetry. Since ΔH varies from step to step in a cycle, we may write $x = f_1(\Delta B_1)/f_2(\Delta H)$, assuming that the change in surface induction is a function of ΔB_s , the change in longitudinal induction. Figure 2 refers to 36% nickel-iron. It appears that the surface induction effect rises to a maximum with $\Delta B_s/\Delta H$ and then falls slightly; the maximum corresponds to the setting of the compensating coil for $H \approx 0$. The circles, triangles and crosses denote measurements made with the

three cycles of magnetization, A, B, and C, with maximum fields of approximately 53, 173, and 350 oersteds, respectively.

§ 3. EXPERIMENTAL RESULTS

The materials used in the present investigation were very kindly supplied, in the hard-drawn state, by Dr. L. B. Pfeil. Their compositions, with the exclusion of that of a nickel-copper-aluminium alloy similar in composition to monel metal, except for the addition of 1.5% of aluminium, are given in the following table. Relevant portions of the largest hysteresis cycles for the several alloys are reproduced in figures 21 and 22.

Table 1

Alloy	Composition per cent						
	Ni	Cu	Fe	C	Si	Mn	S
Monel metal	67.35	30	1.4	0.15	0.1	1.0	0.01
W.5	Bal.	0.1	0.4	0.05	4.0	0.6	0.005
W.6	Bal.	0.1	0.4	0.05	2.5	0.6	0.005
36% nickel-iron	36.0	0.1	Bal.	0.05	0.15	0.5	0.005
42% nickel-iron	42.0	0.1	Bal.	0.05	0.15	0.5	0.005
48% nickel-iron	48.0	0.1	Bal.	0.05	0.15	0.5	0.005
'Mangonic'	Bal.	0.1	0.4	0.05	0.1	3.0	0.005

Monel metal

For this alloy we have the following data:

Cycle	H_m	I_r	H_c	I_m	α	R
B	173	3.6	2.1	9.3	8.0	394
C	365	8.4	2.0	22.5		

In the above list, and in others which follow, H_m is the maximum effective magnetizing field in the solenoid, expressed in oersteds; I_r is the retentivity, or residual magnetization, in gauss; H_c is the coercivity, in oersteds; I_m is the highest intensity of magnetization in a cycle, in gauss; α is the coefficient of linear expansion, in units of $10^{-6} \text{ deg}^{-1} \text{ c.}$; R is the resistance, in microhms per cm. length of the specimen.

In the case of monel metal, the magnetization is so small that demagnetization effects and eddy current effects are negligible. The Q , I and other curves, for the virgin magnetization and for cycle C, are given in figures 3 and 4, respectively. Following the procedure of Paper I, the sum of the heat changes for the individual steps $\sum dQ$ is denoted by Q ; these changes were measured as the effective solenoid field was changed, in the case of closed cycles, from H_m to the same maximum in the opposite sense, and so on. In order to economize in graph space the values, as in figure 4, are plotted for one half-cycle only. In general, the changes in the remaining half-cycle can readily be obtained by rotating the existing Q curve about the axis of ordinates and displacing it vertically until its starting point coincides with the position originally occupied by its end point. In the same figures the values of $\int_{-H_m}^H H dI$ and of $\left(\int_{-H_m}^H H dI \right) - Q = E_T$ are plotted against I .

Warburg's law is confirmed within the limits of experimental error. The energy changes deduced from $\oint H dI$ are so minute that the Q, I curve is practically the same as the curve of internal energy E_T against I , but reversed in sign. The heat changes are large and there is little change in the general form of the curves for cycle A and for cycle B. Comparing the Q, I curves for virgin magnetization with those for the cycles, it is found that the portions of the curves beyond $I=0$ in the latter case are practically the same as for the virgin material. It would seem that practically no thermal changes occur in the virgin material until the reversible 90° boundary displacements are succeeded, beyond the knee of the magnetization curve, by vector rotations.

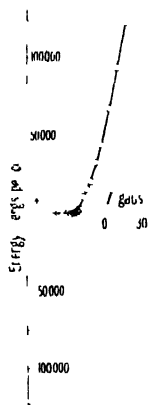


Figure 3. Virgin curve for monel metal. H_m , 365.

— Q, I ;
 ---- $\int H dI, I$;
 $\int H dI - Q, I$.

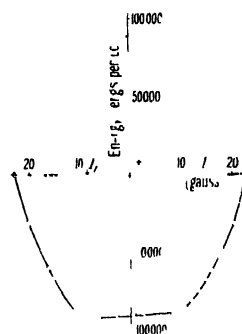


Figure 4. Monel metal. Cycle C. H_m , 365.

— Q, I ;
 ---- $\int H dI, I$;
 $\int H dI - Q, I$.

NOTE.—In all the appropriate curves which follow, the Q, I curves are plotted as full lines, the $\int H dI, I$ curves as broken lines and the $\int H dI - Q, I$ curves as dotted lines, as in figures 3 and 4.

Nickel-Copper-Aluminium

The composition of this alloy is similar to that of monel metal, the main difference being the addition of 1.5% of aluminium. For cycle C the following data, expressed in the usual units, were found.

$$H_m \text{ 350; } I_m \text{ 2.7; } I_r \text{ 0.15; } H_e \text{ 9.5.}$$

With the apparatus working under normal conditions the thermal changes were too small to be recorded with certainty; this is in striking contrast with the results for monel metal. There were undoubtedly indications of slight initial cooling and subsequent slight heating effects between I_m and the knee of the hysteresis curve in the cycle measurements. A slight heating in the case of the virgin material was found beyond the knee.

These results suggest that such phenomena may be of some importance in connection with the measurement of heat losses in transformer materials exposed

to alternating-current fields, as a function of frequency. If there is any time lag between the magnetization processes and those responsible for the comparatively huge evolution and absorption of heat found with such materials as monel metal—and such processes may be present to some extent in all ferromagnetic materials—there is bound to arise a new form of heat loss, hitherto little studied, which must be a function of frequency. We must emphasize that so far we have found no proof of the existence of such time lags.

Nickel-Silicon alloys, W.5 and W.6

The magnetic and other data, expressed in the usual units, are given below:

	Cycle	H_m	I_m	H_c	I_r	α	R
W.5	A	58	17.4	3.40	3.02	10.7	355
	B	173	30.8	4.05	4.10		
	C	370	37.3	4.60	4.56		
W.6	A	58	169	18.8	103	10.0	190
	B	173	264	20.6	122		
	C	365	318	20.2	210		

Demagnetization effects were negligible in the case of W.5; for W.6 the demagnetization factor was assumed to be that of an ellipsoid of revolution with the same length-to-diameter ratio. Eddy current effects were absent. The Q , I and cognate curves for W.5 are shown in figures 5 and 6. The thermal changes are

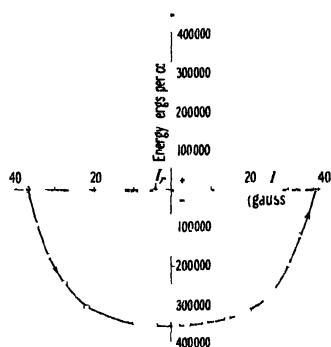


Figure 5. Alloy W.5. Cycle C.
 H_m , 370.

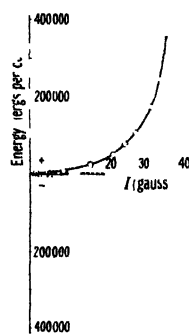
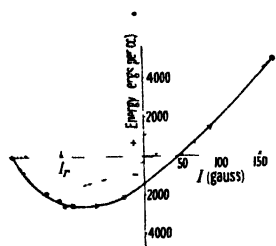
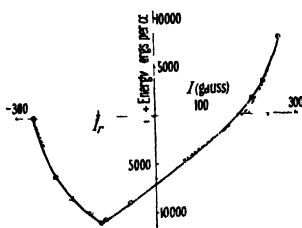
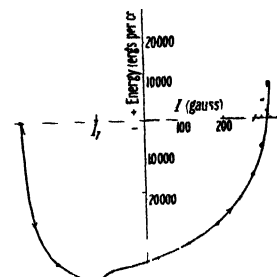


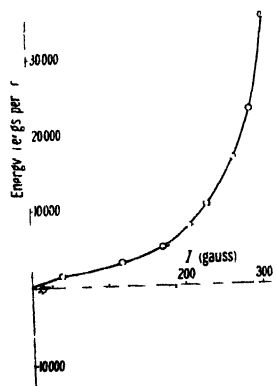
Figure 6. Alloy W.5. Virgin curve.
 H_m , 370.

enormous; once again the E_T curves are practically mirror images of the Q curves. We have had little opportunity of making measurements at different temperatures, and we do not feel that the present apparatus is well designed for the purpose, but some preliminary measurements were made with W.5 by cutting off the solenoid cooling water to cause rises of temperature from 17° c. to 25° c. and to 42° c. At 25° c. the heating effects of cycle B were halved, and at 42° c. reduced to one fiftieth of their original value; more work on these phenomena is clearly necessary.

The Q , I curves for W.6 vary so much with H_m that the results for the three representative cycles as well as the virgin curve are shown in figures 7, 8, 9 and 10.

Figure 7. Alloy W 6. Cycle A.
 H_m , 58.Figure 8. Alloy W 6. Cycle B.
 H_m , 173.Figure 9. Alloy W 6. Cycle C.
 H_m , 365

The Q curve for cycle A is very reminiscent of the corresponding curves for hard-drawn nickel obtained by Bates and Weston, although for W.6 the E_T curve lies above the I axis and its maximum is not at $I = I_r$; but it must be mentioned that the energy scale is so extended that the reliability of the measurements may be over-estimated. The really interesting feature is that as H_m increases, the Q curve becomes more accentuated or "peaked", which is a very unusual phenomenon in our work. Obviously, an abnormal change is taking place at $I = I_r$.

Figure 10. W 6 alloy. Virgin curve
 H_m , 365

Nickel-Iron alloys

The first of these alloys, 36% nickel-iron, approximates in composition to invar, which was examined by Bates and Weston (1940, 1941). The present data, in the usual units, were as follows: the sequence of H_r values with increase in H_m is rather unusual.

ycle	H_m	I_m	H_r	I_r	α	R
A	52	787	4.86	440	2.3	415
B	167	862	4.43	430		
C	362	910	4.04	440		

Loads as great as 4 kgm. produced no measurable Joule cooling, which would indicate a value of α of the order of 0.01×10^{-6} per deg. C., whereas between 20° and 120° C. we found α to be 2.3×10^{-6} by direct experiment. The fluxmeter system could not be calibrated on the basis of Warburg's law, because the values of $\oint H dI$ were so small, and in this instance the alternating-current heating method of calibration was extremely valuable.

Some of the results obtained with 36% nickel-iron are reproduced in figures 11, 12 and 13. The first two figures provide an excellent example of the flattening of the Q , I curves with increase in H_m for measurements between the remanent point and the knee in the reverse quadrant of the I , H curve. The portion of the Cycle C, figure 12, from $I = 0$ to $I = I_m$, is practically identical in form with the virgin Q curve of figure 13. Bates and Weston made measurements

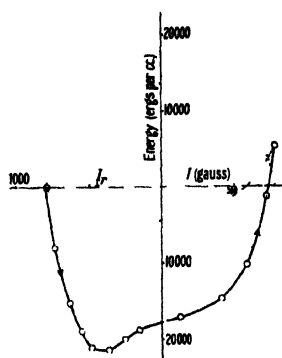


Figure 11 36Ni, 64Fe alloy.
Cycle A. H_m , 52.

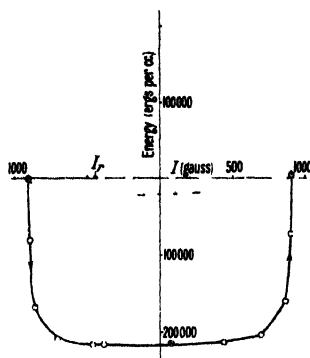


Figure 12 36Ni, 64Fe alloy.
Cycle C. H_m , 362.

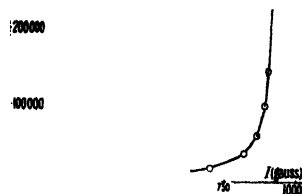


Figure 13. 36Ni, 64Fe alloy. Virgin
curve. H_m , 363

on an alloy of similar composition; they obtained results of the same nature, but the overall range of thermal change in their work was less than one-third of that which would be expected from the present investigation. However, it is well known that slight changes in composition around that of invar have large effects on the coefficient of expansion. The smallness of the latter coefficient for invar has been attributed to the variation with temperature of the spontaneous magnetization giving rise to magnetostriction changes of dimensions which compensate those of ordinary thermal origin during a small temperature change.

The second of these alloys, 42% nickel-iron, gave the following data, in the usual units:

Cycle	H_m	I_m	H_c	I_r	α	R
A	52.3	1017	6.3	697	6.4	482
B	167	1133	6.2	698		
C	363	1157	6.2	690		

Some of the results are shown in figures 14, 15 and 16. An alloy of approximately the same composition was studied by Bates and Weston, but their specimen contained rather more carbon, silicon and manganese than the present one. However, it is satisfactory to note that they obtained the same type and degree of cooling for the intermediate cycle B, the only one which they investigated.

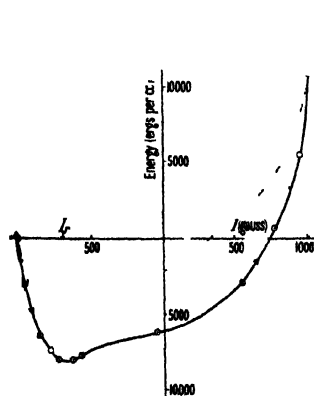


Figure 14. 42Ni, 58Fe alloy.
Cycle A. H_m , 52.3.

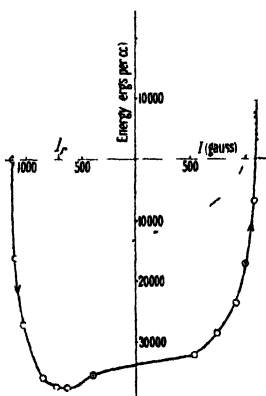


Figure 15. 42Ni, 58Fe alloy.
Cycle B. H_m , 167.

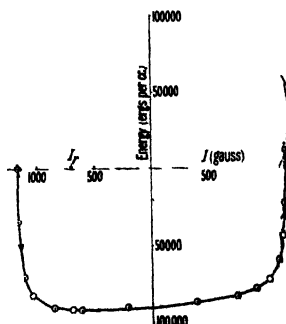


Figure 16. 42Ni, 58Fe alloy.
Cycle C. H_m , 363.

The accompanying figures are reproduced here partly to draw attention to the existence of a peak around I_r , which is flattened out as H_m is increased. The virgin curve is not reproduced, as it is identical in shape, within the limits of experimental error, with that part of figure 16 between $I = 0$ and $I = I_m$. The peak was not found by Bates and Weston; it may be due to the large magnetostriction effects which the alloy is known to exhibit, and which, no doubt, are sensitive to slight differences in composition.

Mangonic

This alloy, consisting of pure nickel alloyed with 3% manganese, gave the following data, in the usual units:

Cycle	H_m	I_m	H_c	I_r	α
A	56.2	307	19.2	232	11.8
B	171.1	418	20.8	267	
C	362.5	466	20.7	269	

The relevant curves are shown in figures 17, 18, 19 and 20. As in the case of W.6,

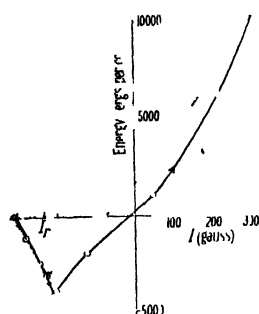


Figure 17. Mangonic.
Cycle A H_m , 56.2.

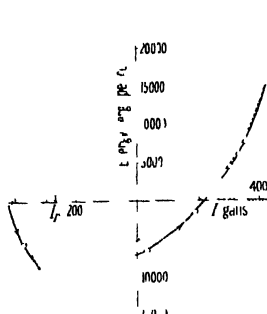


Figure 18. Mangonic
Cycle B. H_m , 171.

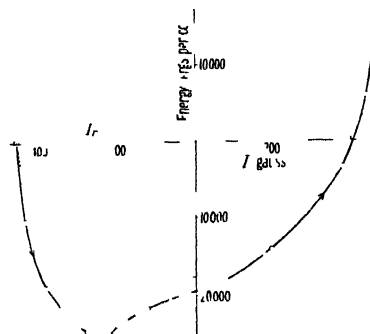


Figure 19. Mangonic. Cycle C
 H_m , 362.

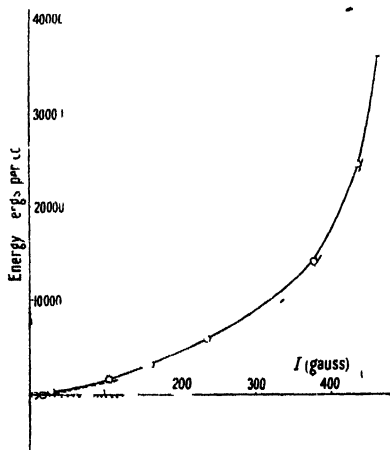
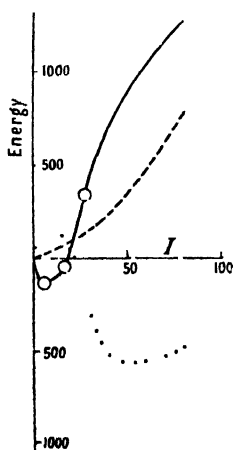


Figure 20. Mangonic. Virgin curve.
(Left-hand curve—portion near origin on enlarged scale.)

a marked peak is found around I_r ; the summit of the peak in the E_T, I curve coincides with the knee of the descending quadrant of the hysteresis curve. With increasing values of H_m the protuberance in the Q curve tends to become more rounded, and, no doubt, it would vanish if we could use sufficiently high values of H_m . For the largest cycle there are no less than three definite peaks in the E_T, I curve.

The virgin curve was difficult to obtain, and this is understandable when the initial portion in the immediate neighbourhood of the origin is plotted on an enlarged energy scale. There must, clearly, be much ambiguity in regard to the orientation of the domain vectors in the demagnetized state. In view of the close connection which one supposes to exist between the permeability of the material in the demagnetized state and that in the region $I = I_r$, similar ambiguity must exist in the latter region. It is thought to be essentially a problem in transverse magnetization, to be tackled along the lines of the suggestions made by Von Engel and Wills (1947) and by Takagi (1939). These amount to the postulate that the distribution of domain vectors in a solid angle $d\omega$ oriented in a direction θ with respect to a direction of easy magnetization in a ferromagnetic crystal obeys an exponential expression of the form $dv_\theta = C \exp(-E/E_0)d\omega$, where E_0 is a quantity whose precise physical significance is at present unknown.

§ 4. DISCUSSION

In conformity with the earlier papers we equate the energy HdI supplied to a ferromagnetic substance in any part of a hysteresis cycle to the sum of the change in energy which manifests itself as an increase in temperature, and which is denoted by dQ in this work, and the change in internal energy dE . We may therefore write

$$HdI = (\partial E / \partial I)_T dI + (\partial E / \partial T)_I dT,$$

or

$$\int (\partial E / \partial I)_T dI = \int HdI - Q,$$

so that the $\left(\int HdI - Q \right), I$ curves allow us to follow the variation of $(\partial E / \partial I)_T$ with I .

The appropriate curves for the measurements with these alloys are given in the preceding figures.

We find that every virgin Q, I curve with the exception of 'mangonic' lies entirely above the I axis, and that there is a marked cooling on reducing the magnetization from I_m to I_r . In the case of large cycles, i.e., large H_m , there appears to be a flattening out of the Q, I curves between I_r and the value of I at the knee of the magnetization curve in the reverse quadrant. We think that this flattening is illusory in most instances, because the larger the cycle the larger are the initial cooling and the final heating effects, and upon these the much slighter departures from linearity, or peaks, such as the one in figure 14, are superimposed. The necessary changes in scale of the several figures also make it difficult to distinguish these departures, and, in any case, it is difficult to measure with accuracy a small quantity superimposed upon a much larger one.

Every E_T, I curve for a closed cycle lies above the I axis, and some of these curves are almost mirror images of the corresponding Q, I curves. It is perhaps advisable to point out that they lie above the I axis because we take the latter as a

datum line, whereas it would be more correct, if much less convenient, to take the horizontal line through the maximum value of E_T on the virgin E_T, I curve.

In all cases the E_T, I curve has a maximum around $I = I_r$. In several instances it is very sharply defined, and we clearly have evidence of a special phenomenon occurring just before the knee of the demagnetization curve. At first sight one might be inclined to associate the phenomenon with that recently described by Hobson (1947), who finds pronounced peaks superimposed in the hysteresis cycles traced on an oscillograph screen. But Hobson used ferromagnetic wires of unstated nature and composition and he has not yet published his experimental arrangements. Consequently, we cannot rule out the possibility that the peaks he observed were due to eddy currents.

We can, of course, say that the heating phenomenon observed in the experiments described in this paper are due to the change in the type of magnetization from what Neel (1944) terms Type III to Type II. This means that at very high fields the domain vectors lie along the direction of the applied field (Type IV); at less high fields they lie along the particular direction of easy magnetization most nearly parallel to the field (Type III); at lower fields they lie along one of the three directions of easy magnetization (Type II), while at very low fields they lie along all directions provided by the axes of the crystal cube (Type I).

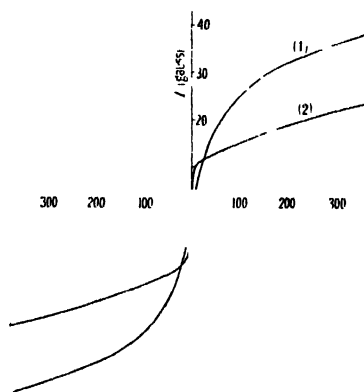


Figure 21. Descending portion of cycle C curve for (1) W.5, (2) Monel metal.

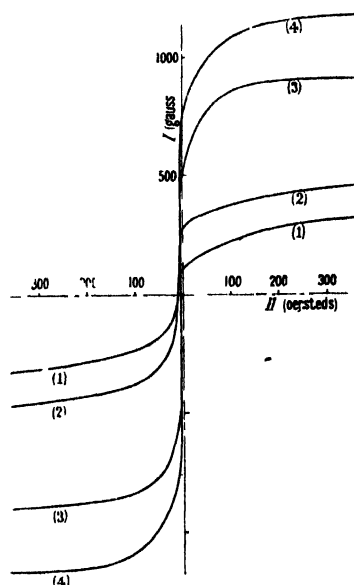


Figure 22. Descending portion of cycle C curve for (1) W.6, (2) Mangonic, (3) 36% Ni-Fe, (4) 42% Ni-Fe.

§ 5. ACKNOWLEDGMENTS

We are deeply indebted to the Government Grant Committee of the Royal Society and to Imperial Chemical Industries Ltd. for the supply to L.F.B. of much of the apparatus used in this investigation. We are similarly indebted to Dr. L. B.

Pfeil of the Development and Research Department of the Mond Nickel Co., Ltd., for the provision of specimens and much helpful advice.

REFERENCES

- BATES, L. F., and WESTON, J. C., 1940, *Nature, Lond.*, **145**, 551.
BATES, L. F., and WESTON, J. C., 1941, *Proc. Phys. Soc.*, **53**, 5.
BATES, L. F., and HEALEY, D. R., 1942, *Proc. Phys. Soc.*, **55**, 188.
BATES, L. F., and EDMONDSON, A. S., 1947, *Proc. Phys. Soc.*, **59**, 329.
HOBSON, P. T., 1947, *Nature, Lond.*, **159**, 436.
NÉEL, L., 1944, *Journ. de Phys. et le Radium*, **5**, 241.
TAKAGI, M., 1939, *Sci. Rep. Jōhoku Univ.*, **28**, 20.
VON ENGEL, A., and WILLS, M. S., 1947, *Proc. Roy. Soc., A*, **188**, 464.

(Note added in proof. By a strange accident in transcription the $\int HdI$, I curve of figure 3 is plotted below, instead of above, the I axis.)

The Adiabatic Temperature Changes Accompanying Magnetization in Low and Moderate Fields : A Further Study of Iron

BY L. F. BATES AND E. G. HARRISON

University College, Nottingham

MS. received 15 August 1947; read 8 November 1947

ABSTRACT. It has been found possible to use a commercial instrument in place of a 'home-made' fluxmeter with the Bates and Weston method of measuring small changes of temperature. The new instrument has been employed to make a further study of the small heat changes which accompany step-by-step changes in the magnetization of iron, using the specimen of annealed Armco iron previously studied by Bates and Healey, and a specimen of H.S. Electrolytic iron (Hilger). In the course of the work the heat changes associated with the virgin magnetization curve have been obtained for the first time.

§ 1. INTRODUCTION

THE heat changes which accompany step-by-step changes in the magnetization of ferromagnetic materials have been studied in some detail by Bates and Weston (1941), Bates and Healey (1943), Bates and Edmondson (1947) and Bates and Harrison (1948). In the present communication we are concerned with measurements on iron, which has proved to be the most difficult metal with which to obtain reliable measurements. Bates and Healey experienced very much trouble from eddy current effects in this material, from strange induction effects and from phenomena which they believed to be associated with a discontinuity in the I, H curve of the kind predicted by Akulov (1931), in addition to the very considerable experimental difficulties inseparable from the measurement of the extremely small changes in temperature, whose order of magnitude may be inferred from a knowledge of the hysteresis cycle. They felt that time-lag effects might also upset their measurements, and they therefore deliberately

sacrificed sensitivity in order to obtain greater stability of zero of the instrument, a special kind of fluxmeter with electromagnetic control, whose deflection recorded the temperature changes.

It occurred to one of us that a 'Tinsley, Type 4789, moving coil ballistic galvanometer, resistance 17 ohms, might well replace the 'home-made' instrument previously used in all this work. Through the kindness of Imperial Chemical Industries, Limited, two of these galvanometers were obtained on loan, and the more suitable for our purpose was immediately adapted as follows. As the permanent magnet in the galvanometer was much too powerful, it was 'shorted' by the bridging of its gap with safety-razor blades, the number of blades being more or less determined by other factors, such as the impedance of the secondary winding on the mu-metal spiral core, referred to below. The galvanometer mirror was of good quality, and, fortunately, was slightly convex with a radius of curvature of the order of 7 metres. The lens fitted in the galvanometer case was therefore removed and replaced by a mica window, covered by a diaphragm with a suitable horizontal slit. This allowed the lamp and scale to be placed about 9 metres from the galvanometer, and thus a sensitivity was obtained some 70% of the maximum attained with the 'home-made' instrument, and equal to about $12 \times 10^{-6}^{\circ}\text{C.}$ per mm. scale deflection. However, this is actually a slight improvement on that used by Bates and Healey; and, in addition, there was much greater stability of zero and a sensitivity which was remarkably constant whatever specimen of ferromagnetic material was under examination, so that optimum conditions could always be relied upon. This enhanced stability was in part due to a new method of attaching the thermojunctions by means of binding thread instead of ebonite collars and grub screws. The sensitivity could, of course, be increased by using even greater scale distance, for the quality of the mirror and the sharpness of the divisions in the field of view would easily permit this to be done, but this would involve taking over another laboratory. The scale divisions could be read to 0.1 mm. with ease.

It was therefore decided to repeat certain of the measurements made on Armco iron, 99.89 per cent pure, by Bates and Healey, and to extend the measurements to a specimen of Hilger H.S. Electrolytic iron, 99.96 per cent pure. The latter is reported to contain 0.02 per cent non-metals of no spectroscopic interest, 0.01 silicon and 0.01 per cent metals other than iron. It is supplied as rods approximately 5 mm. in diameter and 17.5 cm. long. Two of these rods were tapped and screwed together to make a rod 34 cm. long, and the magnetic and thermal measurements were carried out on the middle few centimetres of each part of the composite specimen, thus avoiding the small region around the junction of the two rods.

§ 2. EXPERIMENTAL DETAILS

As in the previous measurements, the specimen rod was mounted vertically along the axis of a water-cooled solenoid which gave a field of 121.7 oersteds per ampere when supplied by a steady current which could be varied in discrete steps. Adiabatic temperature changes of the rod were measured by means of twenty copper-constantan thermocouples. The 'hot' junction of each couple was bound in loose contact with the rod, while the 'cold' junction was very close to, yet thermally insulated from, the rod except for conduction along the material of the couple. Each couple was joined in series with its own primary

winding of low-resistance copper wire wound upon a mu-metal spiral core or ring. A low-resistance secondary coil of 2000 turns was wound upon this core and connected to the ballistic galvanometer. The secondary coil used in the present work could, with advantage, have consisted of many more turns of finer copper wire, for the critical resistance of the unmodified galvanometer circuit was 550 ohms, although it was reduced to 100 ohms by our modifications, while the resistance of the secondary was only about 40 ohms; but we were unable to arrange for another core and winding to be supplied within a reasonable time. Twenty identical thermocouples were used, and as each one was in series with its own primary coil on the spiral core, electrical insulation between the 'hot' junctions and the rod was unnecessary. The close proximity of the junctions minimized thermoelectric currents due to temperature fluctuations inside the solenoid, but the galvanometer showed much tendency to drift when large solenoid currents were supplied; this drift was traced to vibrational movements of the specimen within the heated enclosure.

When the temperature of the specimen was suddenly changed by a small increment ΔT , a current strictly proportional to ΔT flowed in each thermocouple circuit and remained constant for a short interval of time. Each primary coil therefore produced its contribution to the total change in flux in the mu-metal core, and there resulted a deflection of the ballistic galvanometer proportional to ΔT . The arrangement was calibrated by producing standard adiabatic changes of temperature, of the order of 0.001°C. , by the sudden application of a longitudinal stress of F dynes to the specimen, when a fall of temperature $\Delta T_1 = -\alpha TF/QA$ resulted, where α is the coefficient of linear expansion of the specimen, T its absolute temperature, A its area of cross-section and Q the quantity of heat required to raise the temperature of 1 cm^3 of the specimen by 1°C. (Another method of calibration is possible with certain materials, but was not felt to be reliable in the case of iron; see Bates and Harrison 1948.)

The several sources of error peculiar to the work have been fully discussed in the earlier papers, and only the special arrangements made to neutralize the large induction effects need be mentioned here. The latter were not only large but difficult to interpret; for example, without any reason being apparent, it was necessary at times to change the sense of the compensating coil used to neutralize sharp inductive deflections of the galvanometer which otherwise preceded the more leisurely, true, thermal deflections. A single turn of wire had also to be connected around the magnetizing solenoid to act as part of the compensating coil. The reasons for these large inductive effects have never been elucidated. The authors are inclined to believe that they are due to large variations in magnetization and in the electrical conductivity of the surface from place to place, but they may provide evidence that closure domains (Kittel 1946) or *domains de fermeture* (Néel 1944), which are made manifest by the well-known Bitter figures in the case of single crystals of iron etc., are not effective in polycrystalline materials.

It has been frequently emphasized that, if the effects of induction on thermocouples etc. could be sufficiently reduced, they could be completely eliminated by taking half the sum of the galvanometer deflections obtained when a chosen step in magnetization is traversed in the two possible directions. For example, if one described a hysteresis cycle with a maximum range of field from $+H_m$ to $-H_m$ and thence to $+H_m$ oersteds, one would expect to obtain equal galvano-

meter deflections on changing the field from $+H_m$ to $+H_1$ and from $-H_m$ to $-H_1$ oersteds, under ideal conditions with no induction effects present. It was assumed that, if the latter were in the same sense as the thermal effects in the $+H_m$ to $+H_1$ change, they would be in the opposite sense in the $-H_m$ to $-H_1$ change, so that the mean galvanometer deflection would give the true thermal change. Fortunately the authors did not find the marked asymmetry found by Bates and Healey, possibly because a larger choke coil was used in series with the magnetizing solenoid to reduce the rate of change of the magnetic field.

It must again be reported that in the case of iron there is no satisfactory proof of Warburg's law, viz., the energy dissipated in one complete hysteresis cycle is $\oint H dI$ ergs per cm^3 . Eddy current heating played a very large part in these measurements, and there was a large difference between the heat calculated from the magnetic hysteresis cycle data and the nett total heating observed directly when one complete cycle was described. It was assumed that Warburg's law was true and that the above difference was due entirely to eddy current heating. A calculation was therefore made of the sum of the squares of the individual magnetic induction changes of the specimen made during the complete cycle; this was denoted by $\Sigma(\Delta B_1)^2$, and $q \cdot \Sigma(\Delta B_1)^2$ equated to the difference between the thermal change observed directly and that calculated from the magnetic data, q being a constant correcting factor. Thus if a thermal deflection d_1 was observed when a change in induction ΔB_1 was made, the true thermal deflection corrected for eddy current heating was taken to be $d_1 - q \cdot \Delta B_1^2$. The several values of q found for the three main hysteresis cycles performed in this work were equal within the limits of experimental error of about 2 per cent, so that the assumption that it was constant was justified.

The magnetic hysteresis cycle was found by the usual ballistic method using a search coil wound on the specimen and connected to a ballistic galvanometer standardized with a Grassot fluxmeter and a standard mutual inductance. The vertical component of the earth's magnetic field was compensated using a separate winding on the solenoid. Correction was made for the demagnetization effects of the free poles at the ends of the specimen by means of a theory of Becker and Döring (1939) which states that the slope of the virgin curve at the origin, i.e. in the region where the Rayleigh law holds, is equal to

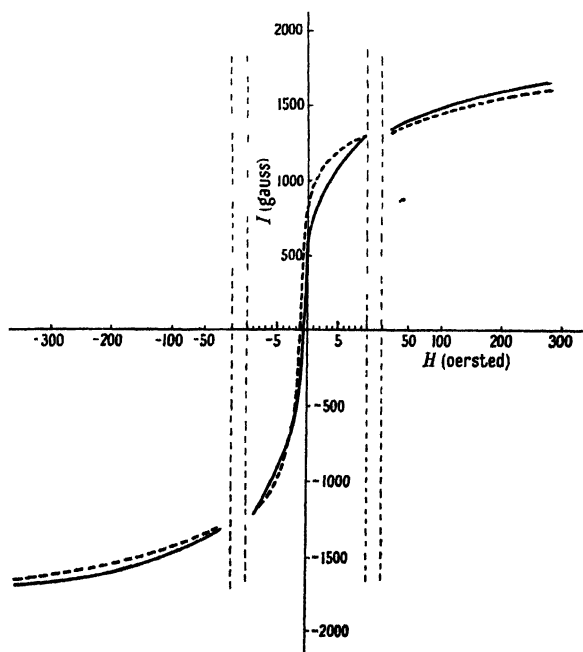


Figure 1. I, H curves.

----- Armco iron; — Hilger H.S. iron.

1/0.364 times that of the I, H curve at the point $I=I_r, H=0$, in the case of annealed iron. If the slope of the former is found to be α , and a line of slope 0.364α is drawn on the uncorrected I, H graph touching the upper, or descending, portion of the hysteresis curve, then the value of H at the point of contact divided by that of I at the same point is equal to the demagnetization factor, because this point must be transferred to the I axis when correction is made for demagnetization. The values of the demagnetization factor so found for the Armco iron specimen were 0.0033 for all three cycles, while for the Hilger iron specimen they were 0.0081, 0.0083 and 0.0086 for cycles A, B and C respectively. The relevant portions of the I, H curves for the larger cycles are given in figure 1. It will be noted that the values of I_r are small in both cases.

§ 3 EXPERIMENTAL RESULTS

Table 1
Armco iron—Magnetic data
for cycle C

$H_{\text{corrected}}$	$H_{\text{corrected}}$	I_{gauss}
373.4	367.9	1650
248.3	243.0	1584
116.1	-111.2	1467
60.9	56.3	1388
16.75	12.50	1288
5.36	-1.80	1070
3.26	-0.53	823
2.07	0.25	668
0.00	0.91	275
2.25	0.98	385
5.24	2.75	732
+ 9.60	+ 5.90	1110
16.9	12.8	1240
- 62.7	58.1	1386
117.3	112.4	1460
+ 250.5	245.2	1585
+ 373.4	367.9	1650

Table 2
Hilger iron—Magnetic data
for cycle C

$H_{\text{corrected}}$	$H_{\text{corrected}}$	I_{gauss}
-369.0	354.4	-1692
-240.3	-227.8	-1449
-111.0	-99.5	-1333
58.5	-47.7	1253
-32.7	-22.5	-1185
15.80	-6.85	1039
9.01	2.11	-802
4.60	0.03	539
1.42	0.62	238
0.00	0.61	-71
+ 1.34	0.61	85
+ 4.63	1.20	+ 399
8.91	2.76	+ 715
16.00	7.54	+ 985
+ 33.0	23.0	+ 1164
+ 59.3	48.6	+ 1242
+ 112.6	101.2	1327
+ 242.0	229.5	1446
+ 369.0	354.4	1692

The more important experimental results are set out in tables 1 and 2 and in the accompanying graphs. The numerical data on which figure 1 is based are given in order that the effect of the demagnetization correction may be fully appreciated. Turning first to the results for annealed Armco iron, which is stated to contain carbon 0.03 %, silicon 0.01 %, manganese 0.03 %, sulphur 0.02 % and phosphorus 0.02 %, it must first be

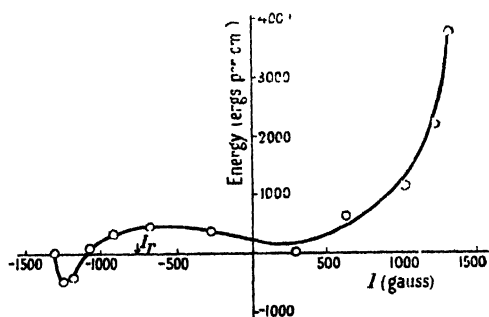


Figure 2. Annealed Armco iron Q, I curve.
 $H_m, 23.9$.

emphasized that the present data are much more extensive than those obtained by Bates and Healey, as results have been obtained for much smaller hysteresis cycles and, for the first time, for magnetization from the virgin state. For example, figures 2 and 3 have no counterpart in the earlier paper. The experimental difficulties of obtaining figure 2 were considerable; this is apparent in the distribution of the experimental points about the curve drawn through them. Both the figures show clearly that it is necessary to go to still lower fields to avoid the initial cooling observed in the early stages of decreasing the field from its maximum value. The 'knee' of the virgin part of the I, H curve lies at about 10 oersteds, and the initial cooling on demagnetization is obviously confined to starting fields greater than this value.

As is usual in this work, only one half of the experimental points are plotted, in order to economize in graph space. In each graph the heat changes ΣdQ are denoted by Q ; these were measured as the effective solenoid field was changed step by step from the stated maximum value

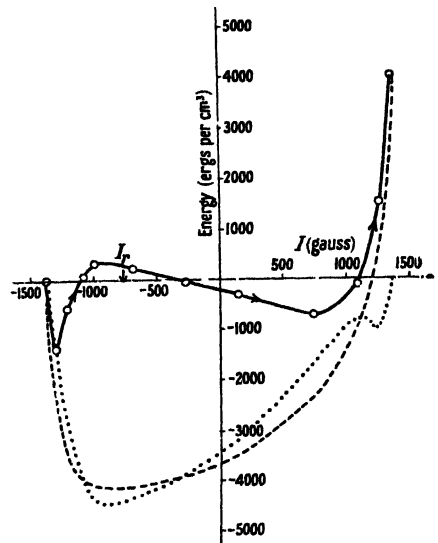


Figure 3. Annealed Armco iron cycle A. H_m , 58.

— — — Q, I ; — — — — $\int H dI, I$;
 $\int H dI - Q, I$.

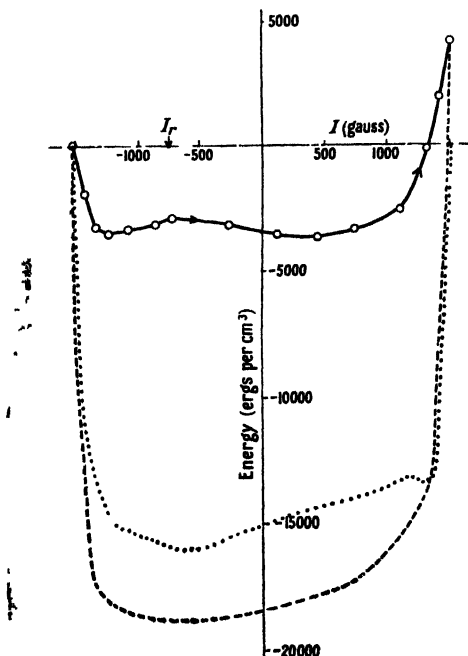


Figure 4. Annealed Armco iron cycle B. H_m , 178.
 — — — Q, I ; — — — — $\int H dI, I$; $\int H dI - Q, I$.

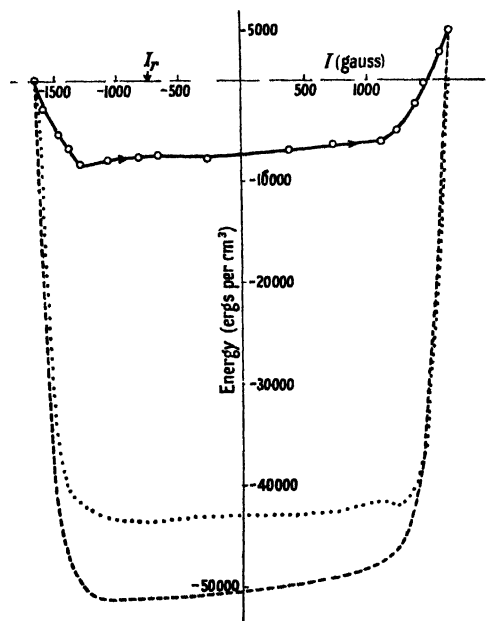


Figure 5. Annealed Armco iron cycle C. H_m , 36.
 — — — Q, I ; — — — — $\int H dI, I$; $\int H dI - Q, I$.

$(-H_m)$ to the same maximum $(+H_m)$ in the opposite sense, and Q is plotted as a function of I . In general, each graph should be completed by rotating the printed Q curve about the axis of ordinates and displacing it vertically until its new starting point coincides with its end point. The values of $\int_{-H_m}^H H dI$ and of $\int_{-H_m}^H H dI - Q$ are also plotted as a function of I .

Figures 4 and 5 should be compared with figures 6 and 7 given by Bates and Healey. On the whole, there is good agreement as far as the general course of the phenomena is concerned, although our experiments lead us to conclude that the Q, I curves always slope upwards to the right between the limits of I equal to I_r and to $-I_r$ on the ordinary I, H curve, in agreement with the finding of Hardy and Quimby (1938). There is, however, definite evidence that the pronounced hump of figure 3 persists to some extent in all the curves, although it tends to be suppressed by large

changes in the scale of the figures in the larger cycles. The $\left(\int H dI - Q\right), I$ curves, later referred to as E_r, I curves, show a kink on the right side only, whereas Bates and Healey obtained traces of a kink on the left side as well, cf. their figure 6.

The results for magnetization of the virgin material shown in figure 6 are very unusual; this is the only example known to the authors of an annealed material which cools as it is magnetized for the first time, starting from the unmagnetized state. There is, of course, a lack of definiteness about the

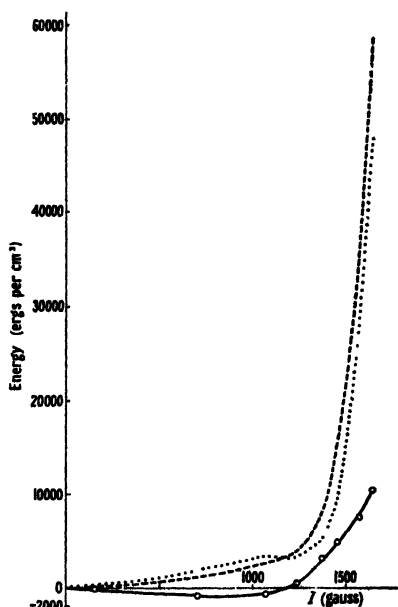


Figure 6. Annealed Armco iron. Virgin curve. $H_m, 370$.

— Q, I ; ---- $\int H dI, I$;
 $\int H dI - Q, I$.

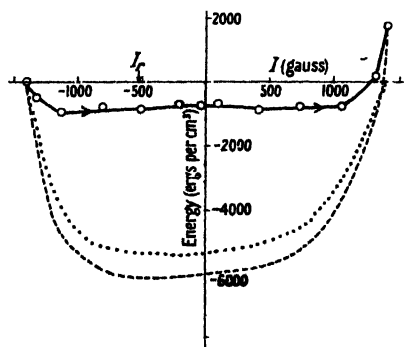


Figure 7. Hilger iron cycle A. $H_m, 57$.
 — Q, I ; ---- $\int H dI, I$; $\int H dI - Q, I$.

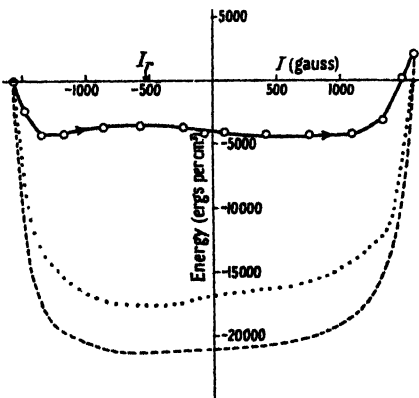


Figure 8. Hilger iron cycle B. $H_m, 178$.
 — Q, I ; ---- $\int H dI, I$; $\int H dI - Q, I$.

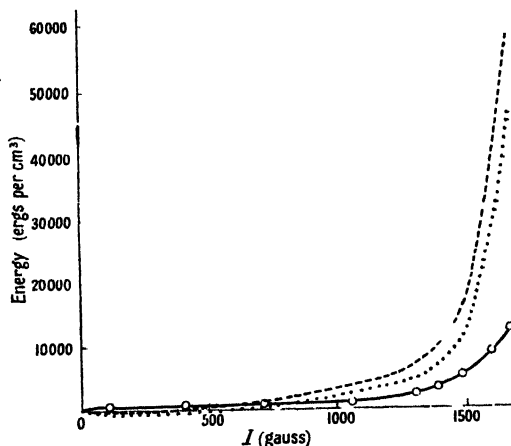
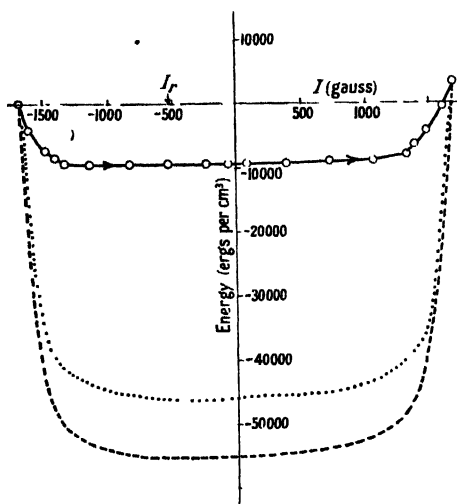


Figure 9. Hilger iron cycle C. H_n , 365. Figure 10. Hilger iron. Virgin curve. H_m , 370.
 — Q, I ; - - - $\int HdI, I$; $\int HdI - Q, I$. — Q, I ; - - - $\int HdI, I$; $\int HdI - Q, I$.

unmagnetized state; the specimen was demagnetized by gradually reducing to zero a large alternating current in the solenoid, with the choke coil removed.

The thermal results for Hilger iron are plotted in figures 7, 8, 9, 10 and 10a; at once a striking contrast between Armco iron and Hilger iron is evident when figures 3 and 7 are compared: there is only a trace of a hump in the latter curve. The Hilger iron was used exactly as received; it was not subjected to any special annealing process by us. The overall heat losses for a complete cycle may be quickly read from the graphs, and, comparing figure 8 with figure 2a given by Bates and Healey for unannealed Armco iron, there is little doubt that the latter iron was magnetically much harder than the Hilger iron. The great difference between the compositions of the two irons clearly lies in the carbon content, and one would naturally seek to explain the hump of figure 3 in terms of the higher carbon content of the Armco iron. Unfortunately, this does not appear to be possible, for we should expect carbon to produce effects similar to those

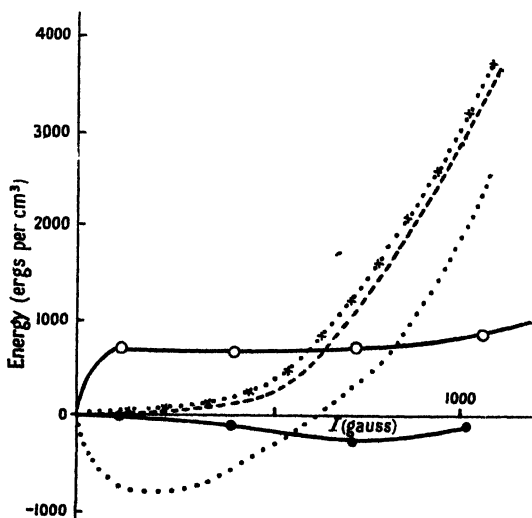


Figure 10a. Hilger iron. Virgin curve showing effects of mode of demagnetization.
 —○— Q, I demagnetized from maximum field of 370 oersted.
 —●— Q, I demagnetized from field of 7.4 oersted.
 - - - $\int HdI, Q$.
 . x . x $\int HdI - Q, I$ for —●— curve.
 $\int HdI - Q, I$ for —○— curve.

reported by Adelsberger (1927) and Constant (1928) in their pioneer work on steels, namely, a more or less continual evolution of heat as the field changes from $+H_m$ to $-H_m$, with a marked evolution of heat as the magnetization changes from $+I_r$ to $-I_r$.

One cannot fail to notice an interesting feature of the hump of figure 3 and of the humps in later figures. If the hump is inverted, it is obvious that the Q, I curve becomes smooth and continuous. It is interesting, too, that if we measure the area enclosed between the hump and a line drawn to touch the lowest portions of the Q, I curves, we find that it is a maximum for cycle A in the case of Armco iron and for cycle B in the case of Hilger iron. It seems fairly safe to presume that for very large values of H_m the hump would completely disappear.

The virgin curve for Hilger iron, shown in figure 10, was obtained after the specimen had been demagnetized by exposing it to a steady field H_m and then to a large alternating field which was gradually reduced to zero, exactly as was done in the case of Armco iron to obtain figure 6. There is no sign of cooling in figure 10, but, again, the unmagnetized state is not uniquely defined, as is seen from figure 10 *a*, where the encircled points of figure 10 are plotted on an enlarged scale for comparison with the points represented by the filled-in circles. The latter points were obtained when the specimen was demagnetized after reaching a field of 7.4 oersteds only, so that the magnetic state was always below the 'knee' of the hysteresis curve. Naturally, the 'filled-in' points were found after one or two preliminary sets of observations had been made and discarded; it is seen that a small cooling effect emerges.

We feel justified in attempting to explain this behaviour on the lines set forth by Néel (1944). He suggests that the I, H curve for a single crystal of iron may be treated in four parts—termed I, II, III and IV, respectively. In part I the crystal is supposed to be magnetically isotropic with the applied field insufficient to overcome the self-demagnetization field, so that within the crystal the field of external origin is neutralized. The domain vectors may then set along six possible directions of easy magnetization, and it may be presumed that the conditions suggested by Von Engel and Wills (1947) also apply. In part II the applied field overcomes the self-demagnetization field, and there is a small resultant field within the crystal acting along a ternary axis, so that the domain vectors can now set along the three directions equally inclined to that axis. In part III the resultant field is large, so that two domain directions only are possible. Finally, when the resultant field becomes very large, we have the state of part IV, in which the vectors are aligned in one direction only. Thus, in part I we are really dealing with a six-phase system as far as the domains are concerned, and it would be truly remarkable if their distributions were the same (*a*) when an ordinary specimen was demagnetized after reaching the state of part IV, and (*b*) after reaching the state of part II at most; indeed, practically any theory of magnetization processes would lead to a similar result.

Néel's treatment leads to the conclusion that, in a bar cut from a single crystal of pure iron parallel to the binary axes and acted on by a field parallel to the axis, the elementary domains are disc-shaped, like fillets perpendicular to the axis. Thus, one may picture a slab cut from a single crystal as made up of transverse slices whose planes are perpendicular to the long axis of the slab, each forming a domain, and magnetized in such a way that there are no free poles inside the material. To avoid the formation of powerful free poles on the lateral surfaces

of the slab, small prism-shaped domains, called closure domains (*domains de fermeture*), abutting upon the edges of the main domains, cause the condition of uniformity of flux to be maintained and close the flux paths of neighbouring discs.

Usually, one tends to think of a single crystal as in no wise unique, and to imagine that a polycrystalline specimen must have magnetic properties resulting very definitely from the superposition of those of a collection of single crystals. But the introduction of the concept of closure domains must cause some modifications of this view, if only because free poles can obviously exist within a polycrystalline specimen. In fact, only in strong fields would one expect to find little or no modification.

§ 4. DISCUSSION

As in the earlier papers, the energy supplied to each cm^3 of specimen in any part of the hysteresis cycle is taken equal to HdI , and assumed responsible for an increase in internal energy dE , according to the equation

$$HdI = (\partial E / \partial I)_T dI + (\partial E / \partial T)_I dT, \quad \text{or} \quad (\partial E / \partial I)_T dI = HdI - dQ,$$

which is correct whether the changes are thermodynamically reversible or not.

In the preceding figures we have therefore plotted the $\left(\int HdI - Q\right), I$ curves, i.e. $\int (\partial E / \partial I)_T dI, I$ curves, whenever possible; such a curve was not included in figure 2, as the magnetic data for such a small cycle were not entirely satisfactory.

The E_T, I curves are much more rounded in the case of the Hilger iron and they do not cross the $\int HdI, I$ curves in the same way as do the corresponding curves for Armco iron. It is thought that this difference can safely be attributed to the greater carbon content of the latter, as evidence of the kind of kink shown in the E_T, I curve of figure 3 was recently found in the case of several alloys (Bates and Harrison 1948). Previous work has led us to expect a minimum in the E_T, I curve in the region of $I = I_r$, referred to the descending portion of the hysteresis curve, and therefore on the left of the I axis in the above figures; this is clearly the case for Hilger iron, but less clearly so for Armco iron, although in neither case is the value of I_r known with high precision.

It is disappointing, to some extent, to have to record that in the present work no confirmatory evidence has been obtained of the strange phenomena observed by Bates and Healey in the range from 350 to 400 oersteds, the region in which Akulov (1931) predicted that a discontinuity ought to exist in the I, H curve for a single crystal of iron magnetized parallel to a $[111]$ direction. Experimental conditions were much more stable than those of Bates and Healey, and the present method of attaching the thermojunctions to the specimen may have been less sensitive to disturbances, or to error, arising from dimensions changes of magnetostrictive origin. For example, with the more rigid fixing system of ebonite collar and lightly adjusted grub screw, alternate expansions and contractions of the cross-section of the specimen might have caused a temporary deterioration in the contact between the 'hot' junction and the specimen which did not occur with the more flexible thread binding.

In conclusion, it appears to the authors that the Q, I behaviour of iron in very low fields depends very much on the purity of the specimen, and that it

would be instructive to repeat the above measurements with different specimens of high purity and known carbon content. They feel that enough is now known of the behaviour of polycrystalline iron and the technique of the method of measuring very small changes of temperature to make the measurement of such changes for a single crystal specimen of iron a possibility. Such an investigation would be extremely interesting because by choosing suitably shaped specimens it should be possible to examine magnetization processes of the type I, III and IV or of the type II only.

ACKNOWLEDGMENTS

Much of the apparatus used in this research was purchased from a grant made to one of us (L. F. B.) by the Government Grant Committee of the Royal Society; the special Tinsley galvanometers were purchased by Imperial Chemical Industries, Limited, for loan to L. F. B.; the specimens of Armco iron were kindly supplied by Professor F. C. Thompson. During a portion of the time spent on this research E. G. H. availed himself of a Government F.E.T. Grant.

REFERENCES

- ADELSBERGER, U., 1927, *Ann. Phys., Lpz.*, **83**, 184.
AKULOV, N., 1931, *Z. Phys.*, **69**, 81.
BATES, L. F., and WESTON, J. C., 1941, *Proc. Phys. Soc.*, **53**, 5.
BATES, L. F., and HEALEY, D. R., 1943, *Proc. Phys. Soc.*, **55**, 188.
BATES, L. F., and EDMONDSON, A. S., 1947, *Proc. Phys. Soc.*, **59**, 329.
BATES, L. F., and HARRISON, E. G., 1948, *Proc. Phys. Soc.*, **60**, 213.
BECKER, R. and DÖRING, W., 1939, *Ferromagnetismus*, p. 160. (Springer, Leipzig.)
CONSTANT, F. W., 1928, *Phys. Rev.*, **32**, 486.
HARDY, T. C., and QUIMBY, L., 1938, *Phys. Rev.*, **54**, 217.
KITTEL, C., 1946, *Phys. Rev.*, **70**, 967.
NÉEL, L., 1944, *J. Phys. Radium*, **5**, 241.
VON ENGEL, A., and WILLS, M. S., 1947, *Proc. Roy. Soc. A*, **188**, 464.

DISCUSSION

Professor W. SUCKSMITH. In addition to the magnetic energy associated with Warburg's law, there will be contributions due to the thermal effect arising from the strain and lattice energies. Have the authors carried out any experiments which would throw light upon the contributions of these effects?

Mr. A. E. DE BARR. With reference to Prof. Bates's remarks about the difficulty of obtaining suitable samples of silicon iron for these measurements, if he will let me know the dimensions I shall be pleased to arrange for specimens to be supplied. These will be rods cut from silicon iron bar before it is rolled into sheets. I should like to ask if the authors have been able to correlate any of their results with measurements of the crystallographic orientation of their specimens. The hard-drawn samples will probably show marked preferred orientation, and this may not remain the same after annealing.

Dr. K. HOSELITZ. The great number of experiments described by Professor Bates and Mr. Harrison show that a variety of types of curves of thermal behaviour accompany magnetization in different materials. I would like to ask whether it has been possible to find a correspondence of adiabatic temperature changes with the various elementary processes during magnetization, such as rotations of domain magnetization and reversible and irreversible boundary movements.

Pure nickel, unstressed and under stress, and permalloy under similar conditions can be studied in this connection.

With regard to pure iron, which, as Professor Bates mentioned, is very sensitive to residual impurities, I wonder whether iron-silicon, the material used in transformer lamination, might prove a more suitable material. It should show very similar characteristics to pure iron, but might not be so sensitive to the presence of small amounts of impurities. Furthermore, as it has a much higher specific resistance, eddy currents should not be so serious and there should not be the necessity for applying such large corrections.

AUTHORS' reply. In the early work many experiments with loaded specimens were made in order to elucidate the effects of strain, and we are making at the moment a survey of the magnetostriction properties. We want, of course, to make experiments on single crystals, and we think we now have sufficient experience to make an attempt.

We shall be most happy to give Mr. De Barr details of our needs. With regard to the question of crystallographic orientation we made experiments on hard-drawn and annealed cobalt in which we obtained very marked differences in thermal behaviour, but, contrary to our expectations, Dr. Lipson found very little difference on x-ray examination.

We have done our best to find the correlation of which Dr. Hoselitz has spoken, and Mr. Harrison pointed out how very similar is the outer portion of the Q, I curve to the outer portion of the virgin Q, I curve; he also drew attention to the 'folding back' phenomenon in the middle of a Q, I curve for a cycle of medium size. We have studied strained and unstrained specimens.

We have not yet tried iron-silicon because we have only obtained this material in the form of thin sheets, and we expect considerable trouble in mounting such specimens; we hope to make an attempt shortly.

The Variation with Magnetization of Young's Modulus for Cobalt

By R. STREET

University College, Nottingham

MS. received 30 May 1947; read 8 November 1947

ABSTRACT. The variation with magnetization of Young's modulus for specimens of annealed and unannealed cobalt has been investigated, for field intensities up to about 700 oersteds, by the method of magnetostrictive oscillation. The decrement of longitudinal oscillation in cobalt rods has also been measured, and found to be of the order of 10^{-3} .

§ 1. INTRODUCTION

IN general the value of Young's modulus of a ferromagnetic material changes when its state of magnetization is varied, and this phenomenon is known as the ' ΔE -effect'. Both static and dynamic methods of measurement have been used to determine Young's modulus for ferromagnetic materials. The static method, in which observations are made on the linear extension of a rod when subjected to an external stress, has been used by Honda and Terada (1907) in their investigation of the ΔE -effect in nickel and other materials. In the dynamic method, resonant longitudinal oscillations are excited in the specimen, which is in the form of a rod, and the value of Young's modulus is determined from observed quantities characteristic of the vibrating system. Various methods have been used for the excitation of the longitudinal oscillations and these are briefly summarized below. Firstly, a combination of cylinders of quartz and the ferromagnetic may be made to oscillate longitudinally by applying an alternating

E.M.F. to suitably disposed electrodes cemented to the quartz. The theory of such composite oscillators has been given by Zacharias (1933), and the method was employed by Siegel and Quimby (1936), Williams, Bozorth and Christensen (1941) and others. Secondly, when a ferromagnetic rod is supported with one end near a small electromagnet, energized by an alternating current, the resultant periodic force excites longitudinal oscillations in the rod. This method was analysed by Akulov (1933) and was used by von Auwers (1933), Wegel and Walther (1935) and Engler (1938). Thirdly, an alternating magnetic field applied to a ferromagnetic rod is generally accompanied by magnetostrictive strains which generate longitudinal oscillations. This method of magnetostrictive oscillation was used in the investigations of the ΔE -effect of Giebe and Blechschmidt (1931) and Yamamoto (1941).

Although the ΔE -effect has been investigated for a number of ferromagnetic materials, detailed measurements of the effect have been made for only two of the three pure ferromagnetic metals, viz. iron and nickel. However, maximum changes in Young's modulus for annealed cobalt subjected to longitudinal magnetic field intensities of 575 and 900 oersteds have been given by Engler (1938) and Yamamoto (1941) as 0.60% and 0.168% respectively. Detailed measurements of the effect in annealed and unannealed cobalt specimens have been made in this laboratory, using the method of magnetostrictive oscillation, and the results are presented here.

§ 2. METHOD AND APPARATUS

If a ferromagnetic rod with zero intensity of magnetization is oscillating longitudinally at its n th harmonic frequency f_0 , then $4f_0^2 l_0^2 = v_0 n^2 E_0$ where l_0 = length of the rod and E_0 , v_0 are the values of Young's modulus and the specific volume (or density⁻¹). The correction term to account for the lateral motion of the rod is assumed to be negligible. If f , l , E , v likewise characterize the vibrating system when the rod has an intensity of magnetization I , and we write $\Delta f = (f - f_0)$ etc., then by differentiation of the above expression we have

$$\Delta E/E_0 = 2\Delta f/f_0 + 2\Delta l/l_0 - \Delta v/v_0.$$

For cobalt the value of $\Delta f/f_0$ is of the order 10^{-3} . $\Delta l/l_0$ and $\Delta v/v_0$ represent the longitudinal and volume magnetostrictive changes for intensity of magnetization I , and have approximate maximum values of 10^{-5} and 10^{-6} , respectively, in the experiments described below. Therefore, for practical purposes $\Delta E/E_0 = 2\Delta f/f_0$, and the relative change in Young's modulus may be measured in terms of the change of resonant frequency of oscillation of the rod.

In this method of measurement the specimen rods of length about 40 cm. and 0.24 cm. in diameter were supported by thin silk threads as close as possible to displacement nodes. Two coils on ebonite formers were placed over, but not touching, the rods at strain antinodes. One coil, referred to below as the exciting coil, was connected to a variable frequency beat frequency oscillator and excited longitudinal oscillations in the rod in the manner already described. The other, detecting, coil was connected to an amplifier and valve voltmeter. Due to the inverse magnetostriction or Villari effect, an E.M.F. was induced in this coil, the magnitude of which indicated the amplitude of oscillations in the rod. Since the final accuracy of the measurements depends on the determination of changes in resonant frequency it is most desirable to have some method of detecting and

eliminating any random changes of the frequency of the current flowing in the exciting coil. A separate cobalt rod with exciting and detecting coils was found to be a convenient form of frequency monitor. This rod was maintained at a constant temperature and its state of magnetization was not altered during an experiment, therefore any apparent changes in its resonant frequency were due to random drifts in the frequency of the beat frequency oscillator. The output of the oscillator may be switched from the rod under test to the monitor when a check of the frequency is required. Alternatively, when the difference of the resonant frequencies of the two rods is large compared with the widths, in frequency, of both resonance curves, the two exciting coils may be joined permanently to the oscillator and both the receiving coils connected in parallel to the amplifier and valve voltmeter. The latter method was usually adopted and a diagram of the apparatus is shown in figure 1. The rods were suspended within an adjustable framework of brass rods. The coils were 2700 turns of 42 s.w.g. enamelled copper wire and their dimensions are given in figure 1. A split disc of copper foil connected to the outer layer of the coil and to earth was cemented to one face of each former and functioned

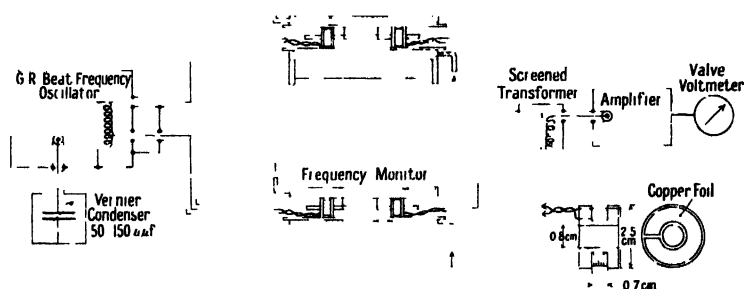


Figure 1. Schematic circuit diagram of apparatus.

as an electrostatic shield. A solenoid, 60 cm. long and wound on a tubular water jacket, surrounded the specimen to be investigated and the frequency monitor rod was permanently magnetized and then placed inside a water jacket. A modified G.R. beat frequency oscillator, Type No. 713, was used in these experiments, the 'cycles increment' condenser being removed and replaced by a Muirhead vernier condenser, Type No. A-429, having a range of 50 to 150 $\mu\mu\text{F}$. The total frequency range of the modified oscillator was 0-37 kc/s. and the change in frequency due to the variation of the vernier condenser was about 2 kc/s. An almost linear relation existed between the changes in frequency and the capacity of the vernier condenser, and since the latter was calibrated in units of $10^{-2} \mu\mu\text{F}$, it was possible to read frequency differences to the nearest 0.1 cycles/sec. At various times during the experiments the frequency controls of the oscillator were calibrated against a G.R. primary frequency standard. The detecting coil was connected to a linear resistance capacity coupled amplifier and diode peak valve voltmeter.

Figure 2 shows a typical resonance curve taken for an annealed cobalt specimen with an applied field intensity of 75 oersteds. The experimental results are in very good agreement with the theoretical curve calculated by assuming a simple resonance phenomenon and that the electrical arrangements adopted obeyed linear laws.

The decrement of the oscillating system under various conditions may be determined from the appropriate resonance curves in the usual way, i.e. decrement $\delta = \pi \Delta f / f$ where Δf = width of the resonance peak at points where the amplitude has decreased to $1/\sqrt{2}$ of its value at resonance, and f = resonant frequency. In order to test the operation of the whole apparatus the decrements for the virgin curve of magnetization and the maximum change in Young's modulus were determined for an annealed nickel rod and the values were compared with those already reported by Siegel and Quimby (1936). The change in Young's modulus was found to be 7.0% (Siegel and Quimby, 6.7%) while the decrements were about 5 % greater than those previously published.

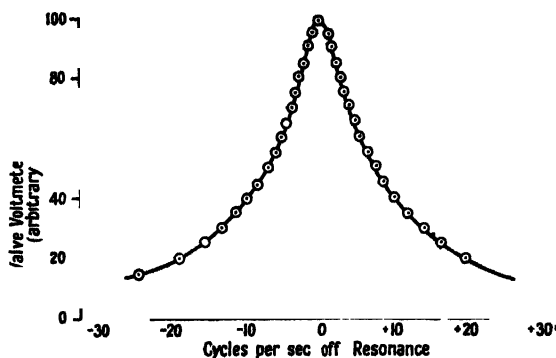


Figure 2. Typical fifth harmonic resonance curve of longitudinal oscillations in an annealed cobalt rod.

§ 3. RESULTS

The measurements were made on specimens of cobalt rod which had previously been used by Bates and Edmondson (1947) in their investigations of the thermal changes accompanying magnetization. The percentage composition was Co 98.40, Ni 0.45, Fe 0.13, CaO 0.23, Mn 0.08, C 0.19, Zn 0.01, Mg 0.11, SiO₂ 0.14, S 0.02, loss in hydrogen 0.24. One specimen of length 39.80 cm. was freshly annealed by maintaining it *in vacuo* for two hours at a temperature of 1000°C., and then allowing to cool to room temperature in 8 hours. The other rod of length 38.13 cm. remained unannealed as received from the manufacturers. Measurements were made on both the specimens using the other as the frequency monitor and the majority of the experiments were carried out at the fifth harmonic frequency of about 30 kc/s.

The voltage induced in the detecting coil was a function of the magnetic condition of the specimen rods and a typical curve illustrating the way in which the detector coil output varied with the applied magnetic field in the case of an annealed cobalt specimen is shown in figure 3. The corresponding curve for unannealed cobalt was similar in form but the maximum induced voltage occurred at a field intensity of 260 oersteds

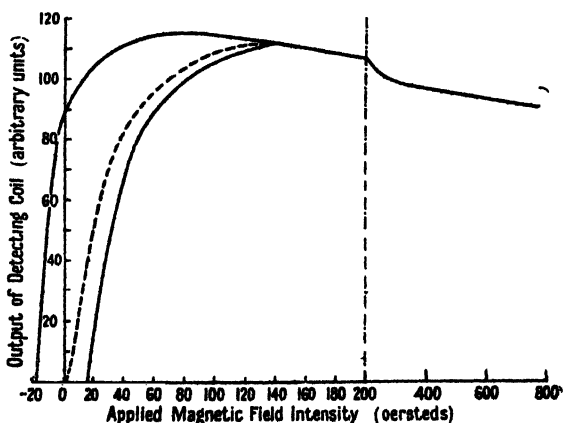


Figure 3. Variation of the output of the detecting coil at resonance with the applied magnetic field intensity.

instead of 70 oersteds. For both specimens the induced voltage was zero when the magnitude of the intensity of magnetization was small, i.e. in the demagnetized state and at the coercive force points on the hysteresis cycle where $I=0$. For this reason it was impossible to investigate the resonance of the rod over the whole range of magnetization available, and this is the main experimental disadvantage of the method. However, in practice the ΔE -effect measurements may be completed by extrapolation without appreciable error. Since the intensity of magnetization along the length of a cylindrical cobalt rod in a uniform externally applied magnetic field is not constant, there will be a corresponding change in Young's modulus for different cross-sections. The resonant frequency of longitudinal oscillation of the rod is determined, to a first approximation, by the elastic modulus of the material near displacement nodes and the inertia of the

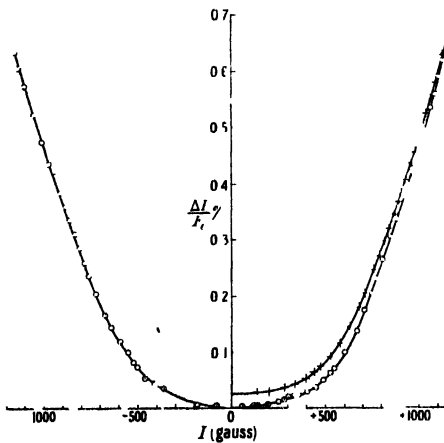


Figure 4. Variation of $\Delta E/E_0$ with I for annealed cobalt.

+ Virgin curve of magnetization.
o = Hysteresis cycle of magnetization.

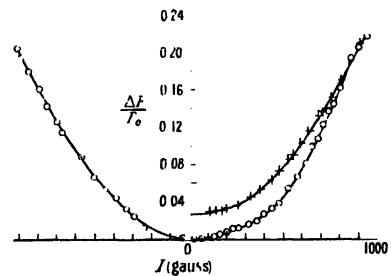


Figure 5. Variation of $\Delta E/E_0$ with I for unannealed cobalt.

+ Virgin curve of magnetization
o = Hysteresis cycle of magnetization.

body in the region of displacement antinodes. Thus by comparison of measurements of the ΔE -effect at different harmonic frequencies, i.e. for different wavelengths of sound in the rod, it is possible to estimate approximately the experimental error due to the above variation of Young's modulus. Results for annealed cobalt taken at the third harmonic frequency agree to within 5% with those determined at fifth harmonic operation. This observation indicates that the error in the measured values of $\Delta E/E_0$ due to the non-uniformity of Young's modulus is small.

The curves given in figures 4 and 5 show the relative changes in Young's modulus for the annealed and unannealed specimens respectively. For both rods the smallest values of Young's modulus are observed at the coercive force points of the hysteresis cycle and the changes in the modulus are expressed in terms of this minimum value, E_0 . For annealed cobalt the present observations agree well with Engler's result, quoted in § 1 above, which corresponds to a cobalt specimen for which I had a maximum value of about 1100 gauss. (Note.—In Engler's work E_0 was taken as the value of Young's modulus for the cobalt in the demagnetized state.) The experimental results for the annealed cobalt taken over the virgin curve of magnetization may be represented to a reasonable degree of accuracy by the

empirical relation $\Delta E/E_0 = 0.023 + 0.67 I^3 - 0.22 I^4$, while the corresponding function for the unannealed specimen is $\Delta E/E_0 = 0.027 + 0.62 I^3 - 0.42 I^4$. In these expressions the intensity of magnetization, I , is measured in kilogauss. In each case the departures of the measured values from the empirical curves are less than $\pm 5\%$.

By extrapolation the fifth harmonic resonant frequencies of the demagnetized annealed and unannealed rods were found to be 30.83 and 30.72 kc/s. respectively. Thus the experimental results together with the value of the density of cobalt, which was found to be 8.71 gm/cm^3 , may be used to determine the absolute values of the moduli of the two specimens. For annealed cobalt $E_0 = 21.0 \times 10^{11} \text{ dynes/cm}^2$, and this figure is in good agreement with the result obtained by Engler (1938). Since frequency differences may be accurately determined with these

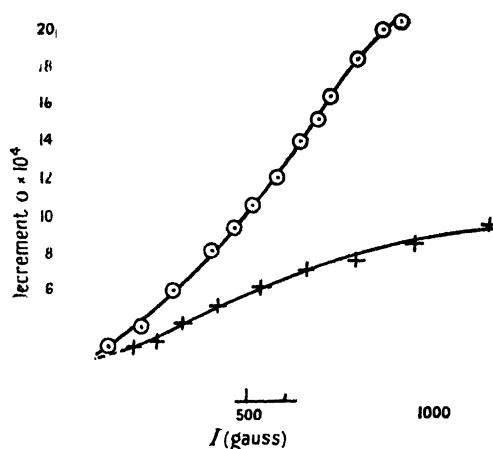


Figure 6. The decrements of longitudinal oscillations in cobalt rods plotted against the intensity of magnetization of the specimens.
+ annealed cobalt. O unannealed cobalt.

experimental arrangements the ratio of the moduli may be accurately calculated from the expression

$$\frac{E_{a0}}{E_{b0}} = \left[\frac{I_a}{I_b} \left(1 + \frac{f_{a0} - f_{b0}}{f_{b0}} \right) \right]^2.$$

In this expression the subscripts a and b refer to the annealed and unannealed rods and the subscript zero indicates extrapolated values representative of the demagnetized state. Hence $E_{a0}/E_{b0} = 1.09$.

The variations of the decrements of oscillations in the two rods for the virgin curves of magnetization are shown in figure 6.

§ 4. DISCUSSION

In the domain theory of ferromagnetism it is postulated that the holomagnetization of a material is the summation of the magnetic effects of elementary domains which are spontaneously magnetized to saturation. In a demagnetized material the magnetization vectors of domains lie along directions of easy magnetization which are determined by the crystalline anisotropy. Two domain processes are supposed to occur during magnetization; firstly, domains having magnetization directions close to the field direction increase in size at the expense of those magnetized away from the field direction and, secondly, rotation of the magnetization

vectors towards the field direction may occur. Various attempts have been made to determine the relative sizes of the domains and the orientations of the magnetization vectors within the domains. These analyses are necessarily statistical in nature and are based on the principle of classical mechanics which states that a system is in stable equilibrium when its energy has a minimum value as, for instance, in the work of Brown (1937) and Takagi (1939). The total energy \mathcal{E} of a domain in a material subjected to an external magnetic field is the resultant of three components, (a) the energy associated with the orientation of the magnetization with respect to the crystal axes within the domain (the magneto-crystalline or anisotropy energy, \mathcal{E}_c), (b) the strain energy arising from the internal stresses of the material and from any externally applied stresses (\mathcal{E}_v), and (c) the magnetic energy of the domain in the field (\mathcal{E}_H).

One consequence of the variation of \mathcal{E}_H due to a change in the external magnetic field is the well known magnetostrictive change in the dimensions of the specimen. Thus, in general, a macroscopic change of dimensions takes place when the characteristic sizes and orientations of the magnetization vectors of the domains are altered. Therefore, when an external stress is applied to a ferromagnetic material having either a positive or a negative magnetostriction coefficient, the resultant strain has two arithmetically additive components: one is of purely elastic origin and the other is due to changes in the characteristics of the domains. The value of the latter component depends upon the relative magnitudes of the total energy \mathcal{E} and the change in \mathcal{E}_H produced by the applied stress. This component will be relatively large when the material is demagnetized, i.e. $\mathcal{E}_H = 0$, and smaller when the material is magnetically saturated, for which state $\mathcal{E}_H > \mathcal{E}_{H\max}$. Hence it is to be expected that the value of Young's modulus of a ferromagnetic in the demagnetized state will be smaller than that for the magnetically saturated material, and this conclusion is verified by the experimental results described above.

The qualitative theory also offers some explanation of the observed difference in the magnitudes of the ΔE -effect for annealed and unannealed cobalt by assuming that in the latter material \mathcal{E} is greater than the corresponding value for the annealed specimen. It seems reasonable to suppose that this difference is due to a larger internal strain energy component in the unannealed cobalt. This analysis of the ΔE -effect has been fully described by Becker and Döring (1939).

The experimental curves show that the Young's modulus of annealed cobalt is about 9% greater than that for the unannealed material. A similar observation has been reported by Cooke (1936) who found that for iron $E_{a0}/E_{b0} = 1.07$. On the other hand, Giebe and Blechschmidt (1931) have shown that the modulus of nickel decreases on annealing. This has been verified by the present method using two nickel rods of diameter 0.40 cm. one of which was annealed by maintaining it at a temperature of 1000° C. for four hours and then allowing it to cool slowly to room temperature. The value of E_{a0}/E_{b0} was 0.90. Experiments are now being undertaken in this laboratory on the ΔE -effect in materials which are progressively strained and it is hoped that the work will give more information on this anomalous behaviour of the three materials.

ACKNOWLEDGMENTS

The author wishes to record his thanks to Messrs. Ericsson Telephones Ltd. for the loan of a suitable beat frequency oscillator and also for their help in the

frequency calibration of this instrument. He is indebted to Professor L. F. Bates for suggesting this problem and for advice and encouragement throughout the work.

REFERENCES

- AKI'LOV, N., 1933, *Z. Phys.*, **85**, 661.
 VON AUWERS, O., 1933, *Ann. Phys., Lpz.*, **17**, 83.
 BATES, L. F., and EDMONDSON, A. S., 1947, *Proc. Phys. Soc.*, **59**, 329.
 BECKER, R., and DÖRING, W., 1939, *Ferromagnetismus*, pp. 336-357. (Springer, Berlin.)
 BROWN, W. F., 1937, *Phys. Rev.*, **52**, 325.
 COOKE, W. T., 1936, *Phys. Rev.*, **50**, 1158.
 ENGLER, O., 1938, *Ann. Phys., Lpz.*, **31**, 145.
 GIEBE, E., and BLECHSCHMIDT, E., 1931, *Ann. Phys., Lpz.*, **11**, 905.
 HONDA, K., and TERADA, T., 1907, *Phil. Mag.*, **13**, 36.
 SIEGEL, S., and QUIMBY, S. L., 1936, *Phys. Rev.*, **49**, 663.
 TAKAGI, M., 1939, *Sci. Rep. Tōhoku Imp. Univ.*, **28**, 20.
 WEGEL, R. L., and WALTHER, H., 1935, *Physics*, **6**, 141.
 WILLIAMS, H. J., BOZORTH, R. M., and CHRISTENSEN, H., 1941, *Phys. Rev.*, **59**, 1005.
 YAMAMOTO, M., 1941, *Phys. Rev.*, **59**, 768.
 ZACHARIAS, J., 1933, *Phys. Rev.*, **44**, 116.

DISCUSSION

Professor L. F. BATES. It was hoped that these very interesting experiments by Mr. Street would help in the interpretation of the kind of experimental results which Mr. Harrison and I described earlier today. We are very much puzzled by the failure of a ferromagnetic rod to oscillate even when it possesses marked magnetization. As Mr. Street has pointed out, we have only a tentative suggestion by Snoek to provide any kind of explanation of this failure to oscillate, and, if that explanation fails to satisfy us, we are completely in the dark. I therefore hope that if any Fellows present today have any ideas on this subject they will let us hear them.

Professor W. SUCKSMITH. Has an x-ray investigation been made of the specimen on which data have been given? Cobalt is a mixture of face-centred and hexagonal lattices, and one would expect the two different structures to make different contributions to the modulus.

AUTHOR'S reply. An x-ray analysis of the specimens of annealed and unannealed cobalt was very kindly made by Dr. Lipson, and this showed that there was little difference in the crystalline structure of the two materials. A summary of the results of Dr. Lipson's analysis is given in a note added to the paper by Bates and Edmondson cited in the list of references above.

The Electric Strength of Dielectric Films

BY K. W. PLESSNER

Electrical Engineering Dept., Imperial College of Science and Technology

Communicated by Willis Jackson ; MS. received 17 June 1947

ABSTRACT. An attempt is made to verify the rise in electric strength predicted by Fröhlich's theory as the thickness of the dielectric specimen approaches the electronic mean free path. The results obtained on a number of materials both crystalline and amorphous are given and discussed in relation to this theory.

§ 1. INTRODUCTION

THE theory of electric strength which appears capable of the most reliable predictions is that developed by Fröhlich (1937, 1939, 1941). It has shown quantitative agreement with experiment for ionic crystals such as the alkali halides and mica, and qualitative agreement has also been obtained for mixed crystals. Fröhlich's theory is to a large extent based on the calculation of a time of relaxation of the electrons and, associated with this, an electronic mean free path in the dielectric. As the thickness of the dielectric specimen decreases and approaches the dimension of the mean free path, an increase in electric strength is predicted, and this has, in fact, been observed by Austen (Austen and Hackett 1939, Austen and Whitehead 1940) in measurements on thin films of mica. These tests form the only published verification of this particular aspect of Fröhlich's theory, and it is the object of the present work to extend the measurements to other materials having different values of the mean free path.

Calcium fluoride, sodium fluoride and potassium bromide provide a convenient range of mean free paths and were selected for the tests. For comparison, two amorphous materials, silica and polystyrene, for which no dependence of the electric strength on thickness was expected from Fröhlich's theory, were also tested.

§ 2. THEORY OF ELECTRIC STRENGTH

According to Fröhlich, the breakdown field strength is to be calculated from the condition that an electron should gain more energy from the field than it gives up to the crystal lattice in the form of heat. Both these energy components involve τ , the average time between collisions (time of relaxation), and in the final expression for the electric strength $1/\tau$ enters as a factor, showing that the strength increases as the time of relaxation decreases, i.e. the shorter the mean free path, l . Thus, the less ordered the structure the higher the electric strength, since the mean free path is shorter. The increased strength of mixed crystals over that of the pure components observed by von Hippel (1937) is an example of this effect of disorder, and when other things are equal, as, for example, with quartz and fused silica, the electric strength of the amorphous phase is higher than that of the crystalline phase. From similar reasoning it is predicted that in a crystalline chain compound, such as a hydrocarbon, the electric strength in a direction across the chains should be higher than along the chains.

Fröhlich gives the following formula for the mean free path, applicable to diatomic ionic crystals:

$$l = \frac{16E^2a^2}{2^{\frac{2}{3}}\pi^2(\epsilon - \epsilon_0)e^2h\nu} \left/ \left(1 + \frac{2}{e^{h\nu/kT} - 1} \right) \right.,$$

where

E = energy of the electron, to be equated to J , the internal excitation energy of the lattice,

ν = reststrahlen frequency in the infra-red,

a = half-lattice spacing,

ϵ, ϵ_0 = dielectric constants for static and visible frequency fields, respectively.

e, h and k have their usual meanings.

A slightly modified expression is used for more complex crystals.

The values of l given in table 1 for a number of substances have been calculated from this formula using infra-red data given by von Hippel (1937) and values of the internal excitation energy provided by Hilsch and Pohl (1930).

Table 1

Substance	Mica	Crystal quartz	CaF ₂	NaF	KBr
Mean free path in 10 ⁻⁸ cm. ..	0.5	1.2	1.5	1.1	2.5

If the behaviour of the halides is analogous to that of mica, then, according to the work already quoted (Austen *et al.* 1939, 1940), a rise in electric strength should be noticeable at thicknesses between 10⁻⁴ and 10⁻⁵ cm. The mean free path in amorphous materials cannot at present be calculated, but it should be appreciably shorter than for crystalline ones. Thus, no dependence of the electric strength on thickness would be expected for amorphous silica for thicknesses of this order.

In von Hippel's view of the mechanism of breakdown no mean free path is defined, his criterion for the limiting field being that the electrons should gain about twice the 'reststrahlen' quantum of energy ($h\nu$) in traversing the distance between neighbouring ions, i.e. a distance of only about 3×10^{-8} cm. However, both von Hippel's and Fröhlich's points of view require that the electrons be accelerated from thermal energies to the internal excitation energy J of the lattice. This implies that the voltage necessary for breakdown cannot be less than J expressed in electron volts. A minimum is therefore set to the electric strength equal to J/t v/cm., J being of the order of 10 v.; this minimum becomes larger than the normal electric strength at thicknesses t between 10⁻⁵ and 10⁻⁶ cm.

§ 3. EXPERIMENTAL METHOD

The first problem in the work undertaken was the production of films suitable for breakdown tests, of thickness appropriate for showing the effects predicted and of a texture which would not adversely influence the processes postulated in the theory. Thus, single crystal sheets of thickness between 0.1 and 10×10^{-5} cm. would be ideal, but in practice could hardly be prepared to be sufficiently uniform.

The method adopted was therefore that of evaporation under high vacuum, the vapour condensing on a suitable surface to form a uniform thin film. The process is used extensively by the optical industry in the 'blooming' of lenses to produce non-reflecting surfaces, and is particularly suitable for giving uniform films of any required thickness up to about 10⁻⁴ cm.

Here, the films were condensed on a layer of silver on a glass slide, the silver having been newly deposited by the same process of evaporation. This acted as the basic electrode for the electrical tests, and also enabled an approximate determination of the refractive index of the film to be made. A further layer of silver was deposited over part of the dielectric film to allow the film thickness to be measured by interferometry.

(i) The preparation of films

For the preparation of vapour-deposited films an evaporation plant of the usual design has been employed, consisting of bell jar and base-plate mounted on an oil-diffusion pump, so that a vacuum of about 10⁻⁵ mm. Hg could be attained.

Heavy electrodes were brought through the base-plate to supply the evaporation heaters, and a brass cone fitting into the top of the bell jar allowed a 'clean-up' discharge to be passed through the vessel. This cone could be rotated to move a revolving carrier inside the vacuum, so bringing the glass slide used as substrate over the various shields and filaments. Figure 1 shows the arrangement inside the bell jar, with a clean glass slide in position to be silvered. Two filaments of molybdenum strip were employed, one to evaporate silver, the other for evaporating the crystalline powders of CaF_2 , NaF or KBr . Materials of the purest grade commercially available were used, and weighed amounts were evaporated completely at temperatures ranging from 730°C . to 1360°C . In the preparation of silica films, the filaments were of tungsten wire (m.p. 3400°C .), since, although molybdenum does not melt below 2600°C ., it was feared that a small amount might vaporize at the high temperature ($>1700^\circ\text{C}$.) required to evaporate silica.

Since the mean free path in air at 10^{-5} mm. Hg is of the order of 8 metres, the evaporated molecules travel essentially in straight lines and behave to some extent like rays of light. Thus, sharp shadows can be thrown by placing shields close to the receiving surface, and the intensity of the beam varies inversely as the square of the distance from the source. From the latter law, the thickness of the film could be predetermined approximately, and it could also be estimated that the film thickness should be uniform to within 3% over a two-inch diameter circle of the specimen slide.

Evaporation was normally carried out at a pressure of 10^{-5} mm. Hg as measured by an ionization gauge, and proceeded as follows:—

- (a) The whole of the specimen slide was silvered to opaqueness.
- (b) The carrier was rotated to bring the slide over the second filament, and in this position a narrow strip of the slide was screened off by a piece of brass close to it, as shown in figure 2. The dielectric powder was then evaporated during a period of about 1 minute.
- (c) The slide was moved over a different shield and a small region, including part of the strip left free from dielectric, was resilvered to allow thickness measurement by the method described below.

In the preliminary measurements the specimen was then taken out and tested, but later the breakdown measurements were made with the specimen remaining in the bell jar, using a magnetically operated electrode.

In addition to the vapour-deposited films, some films of polystyrene, an organic polymer, were produced by the evaporation of a dilute solution. Silvered glass slides were immersed in the solution and withdrawn at a constant rate, during which process the solvent evaporated, leaving behind a uniform film of the polymer. By varying the concentration of the solution and the speed of withdrawal, a thickness range from 0.3 to 3×10^{-5} cm. was covered.

(ii) *Electric strength tests*

Electric strength tests were carried out by pressing a highly polished steel sphere $1/16''$ diameter lightly on to the dielectric surface and applying a uniformly increasing voltage between this electrode and the basic silver film until breakdown occurred. As time effects were observed with some materials, a pulse generator was built, the circuit of which is given in figure 3 to the left of

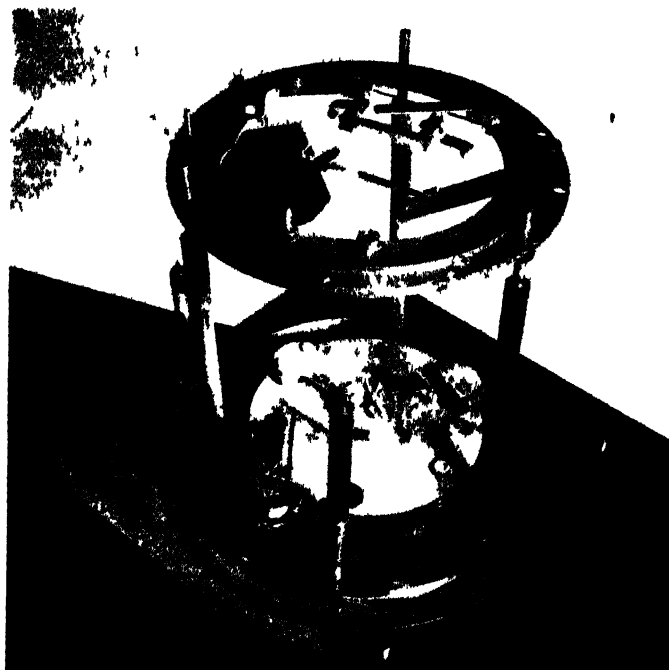


Figure 1

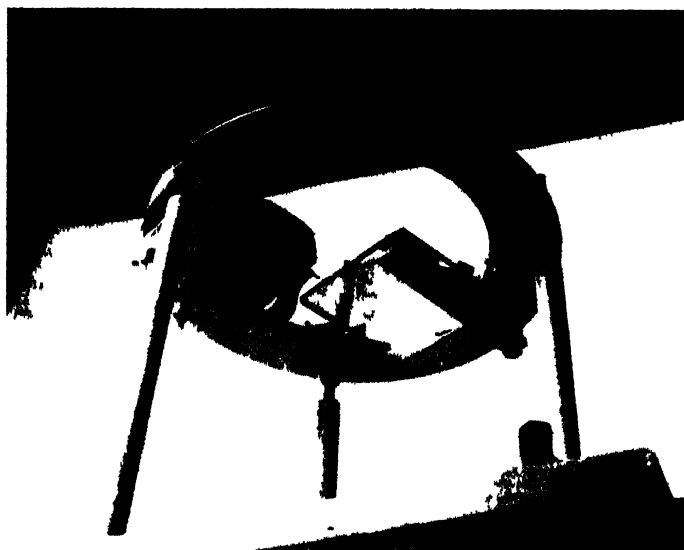


Figure 2

the dotted line. A single pulse, of the triangular shape shown in figure 3 (a), is produced when button A is pressed. The action of the circuit is based on two condensers charging up through two resistances. The first condenser, C_1 , is normally shorted by a triode V_2 , but by pressing button A, the grid of V_2 is connected to a negative potential through condenser C_2 and so C_1 is allowed to charge up through R_1 to give the front of the pulse. At the same time C_2 charges up through R_2 , and when the grid of V_2 returns to zero potential, the valve again becomes conducting and C_1 is short-circuited abruptly, so terminating the pulse. The voltage across C_1 is taken through the cathode follower V_3 and the output is obtained at P from the potentiometer R_3 . Potentiometer R_4 provides a bias

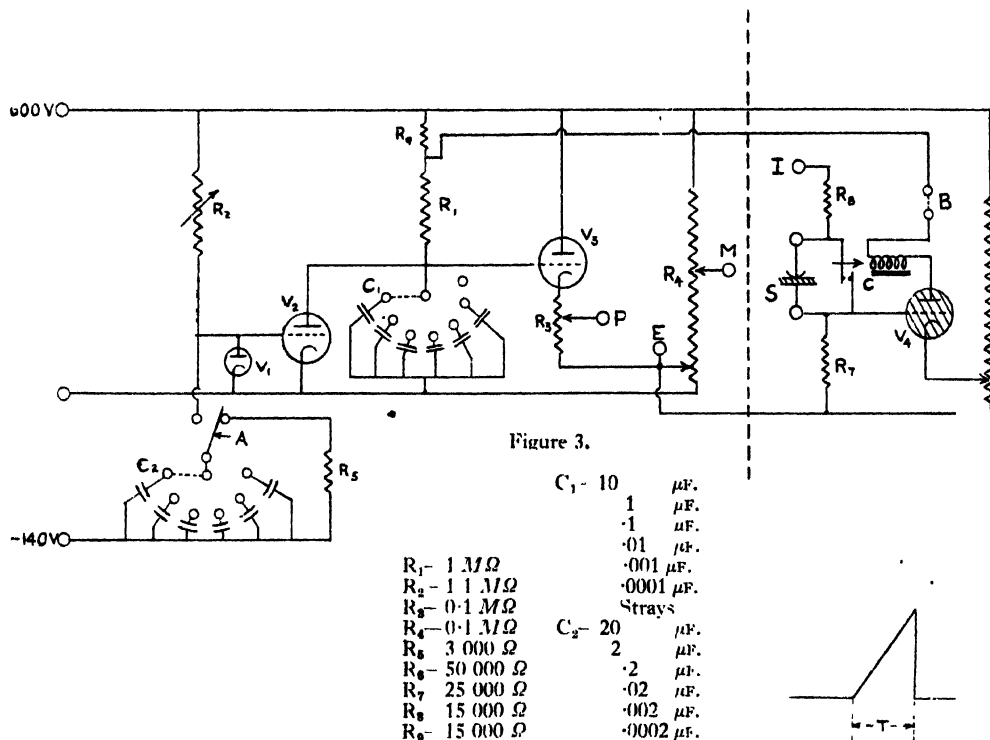


Figure 3 (a).

for V_3 and, at the terminal M, allows manual voltage control, should a very slow rate of rise be required. The pulse duration T could be varied between 10 seconds and 40 μ sec. by controlling C_1 , C_2 and R_2 . The pulse amplitude is variable from 0 to 430 volts.

A number of pulses with gradually increasing amplitude had to be applied until breakdown took place, but once the necessary value was known approximately only a few pulses, starting with an amplitude about 10% below the expected value, were required. The amplitude of the pulses was measured to an accuracy of about 3% on a cathode-ray tube.*

The occurrence of breakdown was indicated by the relay circuit shown to the right of the dotted line in figure 3. The test voltage is connected at I, and is

* It should be mentioned that Turner and Lewis (1947) have used a pulse method for measuring the dependence on field strength of the conductivity of thin glass films.

applied to the specimen S through the safety resistance R_8 and the triggering resistance R_7 . V_4 is a gas-filled triode whose cathode bias is just below the critical value, so that the voltage drop across R_7 , which occurs when the specimen breaks down, causes the valve to strike. The resulting current operates a relay C which short-circuits the specimen to avoid pitting of the electrode, and gives an audible click as an indication of breakdown.

The state of cleanliness and polish of the sphere electrode proved important. When an abnormally high value of the electric strength had been observed, a dust particle or a chip of the film was nearly always found on the electrode. In the early tests, when the specimens were withdrawn from the evaporating chamber and tested in the open, frequent repolishing of the electrode was necessary, until no appreciable marks were visible under a microscope ($\times 120$). In the later tests under vacuum only occasional repolishing was called for.

In the tests carried out in the open, a stream of dry air was blown over the surface of the specimen. This was found insufficient, however, to eliminate

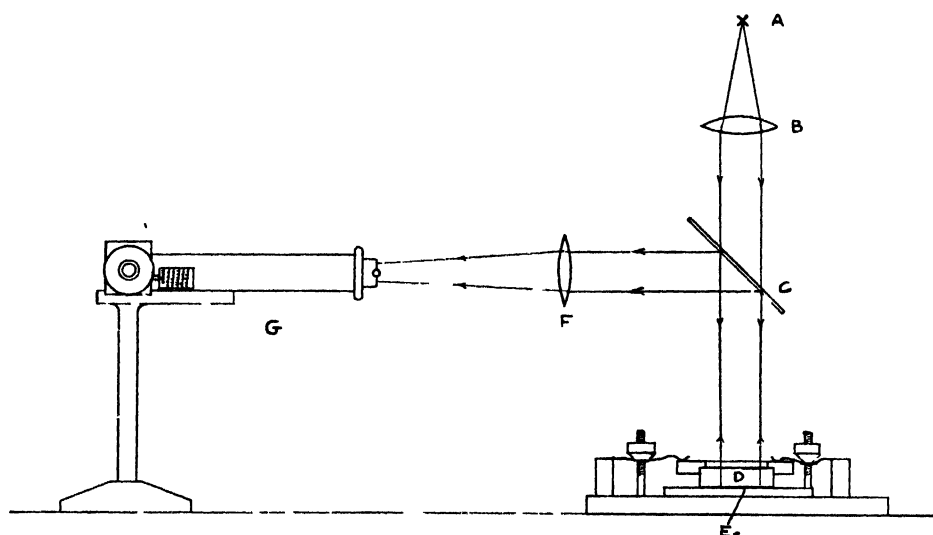


Figure 4.

the effects of moisture, and as stated above, the later measurements were made in the evaporation chamber without exposing the specimen to atmospheric conditions. The effect of moisture was very pronounced, an increase by a factor of 2.5 in electric strength being obtained by testing NaF under vacuum or in thoroughly dry air rather than in the atmosphere.

(iii) Thickness measurement

The thickness of the films was determined by the technique of multiple beam interferometry (Tolansky 1945).

Interference took place between the silvered surface of the specimen slide and the half-silvered surface of an optical flat which was brought into close contact with it. The experimental arrangement is shown diagrammatically in figure 4. White light from the source A is collimated by the lens B and illuminates the interference surfaces consisting of the specimen slide E and the optical flat D pressed together by three adjustable springs. The reflected light from D-E

is then reflected in the half-silvered mirror C, and is focused on to the slit of the spectrometer G by the lens F. In the spectrum sharp dark fringes are visible, corresponding to wavelengths λ , which satisfy the equation $n\lambda = 2t_0$, where n is an integer and t_0 is the optical distance between the silver surfaces.

If t_0 changes by an amount δt_0 owing to a step on one of the surfaces, such as that produced by the shielding of a strip during deposition of the dielectric film, then the fringe system is displaced, as sketched in figure 5.

A shift of one fringe separation corresponds to $\delta t_0 = \lambda/2$, while a relative shift s/d (figure 5) means $\delta t_0 = (s/d)(\lambda/2)$.

The accuracy with which the fringe displacement, and hence the value of δt_0 can be determined, depends on the sharpness of the fringes and on irregularities in the surface of the specimen slide, which impose small ripples on the fringes. Taking into account these factors, and the smoothing effect due to using the mean of a number of readings, it was estimated that the error in δt_0 should not be larger than $\pm 25 \text{ \AA}$.

The step δt_0 is equal to the film thickness t in the region of the specimen slide which was resilvered after depositing the dielectric; whereas in the part with only the first layer of silver, the change in optical thickness is from a film of refractive index μ to air, i.e. $\delta t_0 = (\mu - 1)t$. Thus, both the thickness and the refractive index of the films can be determined by this method.

Although no inaccuracies other than the reading errors mentioned above affect the measurement of film thickness, disturbances due to the imperfect reflectivity of the silver, and to the inadequacy of the equation $n\lambda = 2t_0 + 2\mu t$ to describe interference in a composite gap, influence the determination of the refractive index (Plessner 1946). An approximate correction is possible for thick films, but for films thinner than about $3 \times 10^{-5} \text{ cm}$. the results are not trustworthy.

§ 4. EXPERIMENTAL RESULTS

(i) Vacuum evaporated films

The results of measurements on silica, CaF_2 , NaF and KBr , carried out under vacuum soon after deposition of the films, are given in this section.

The value of electric strength at each value of the pulse duration is the mean of about six readings whose scatter amounted to approximately $\pm 6\%$ for the crystalline films. For silica, where the electric strength was nearly independent of pulse duration, rather more scatter was encountered, and more readings were taken.

Allowing for the scatter, the error in voltage measurement, and the error in thickness measurement, the standard error in the means has been estimated as $\pm 4.5\%$ for film thicknesses greater than 10^{-5} cm . For thinner films the inaccuracies represent a higher percentage, approximately $\pm 10\%$ at $5 \times 10^{-6} \text{ cm}$. and $\pm 20\%$ at 10^{-6} cm .

The values of electric strength found for films of silica are shown in figure 6 plotted against thickness. A slight dependence of the strength on pulse duration was found, but, as its order of magnitude was the same as, or smaller than, that of the scatter, the two have been combined to give the rather extended vertical

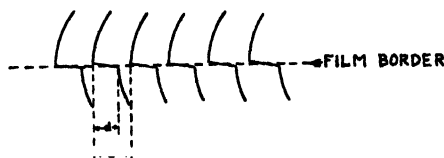


Figure 5.

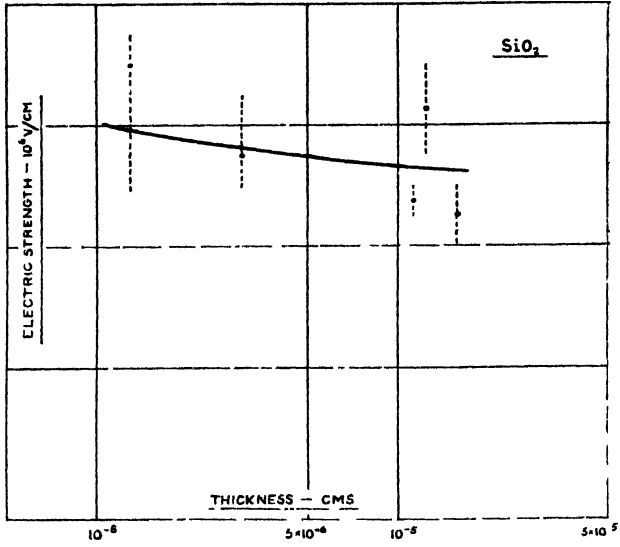


Figure 6.

lines in figure 6. It is seen that there is no significant variation of electric strength with thickness; the slightly higher value found for the thinnest films is not reliable in view of the large error at that thickness.

The three halides CaF_2 , NaF and KBr each show a variation of electric strength with both pulse duration and thickness, as shown in figures 7, 8 and 9, where electric strength is plotted against pulse duration, for a number of constant thicknesses. It will be seen that at a constant pulse duration the strength increases in general with decreasing thickness, as predicted by Fröhlich's theory.

(ii) *Polystyrene films*

The electric strength of polystyrene films was found to be independent of pulse duration, and no significant variation with film thickness between 0.37 and 3×10^{-5} cm. was encountered (table 2).

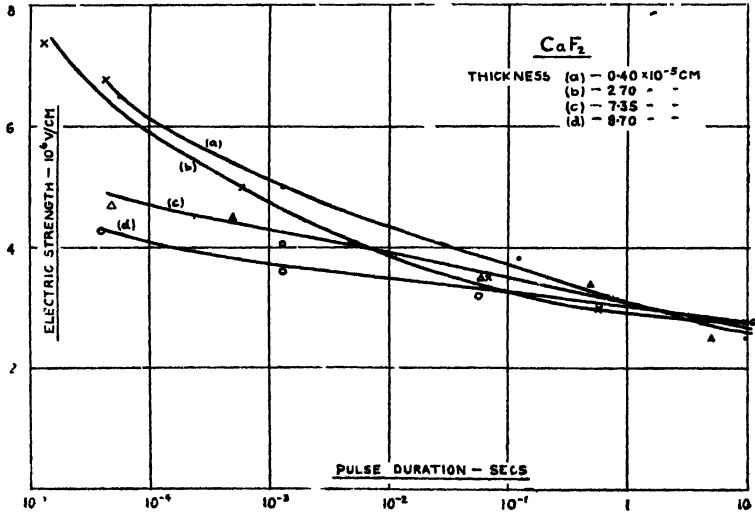


Figure 7.

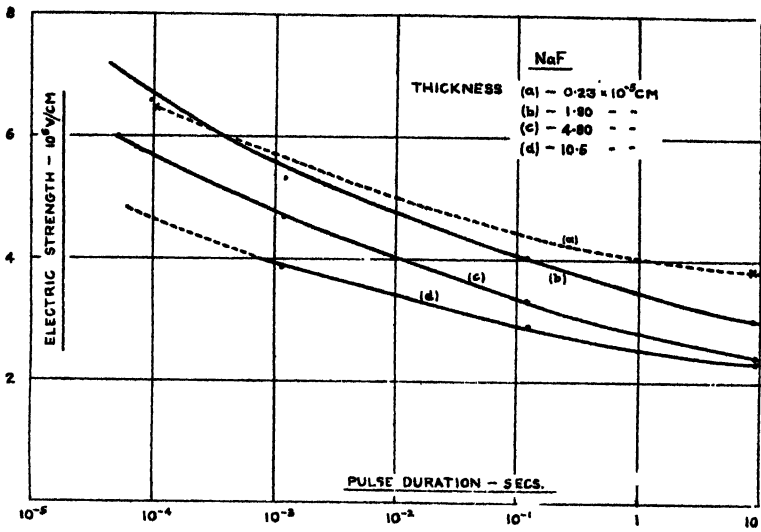


Figure 8.

Table 2

Thickness (10^{-5} cm.)	0.37	0.47	0.84	1.72	3.1
Electric strength (10^6 v/cm.)	4.9	6.0	5.5	5.5	5.6

The figures given are each the mean of about 20 readings, the scatter among which amounted to about $\pm 15\%$.

The overall mean value is $5.5 \times 10^6 \text{ v/cm.}$

§ 5. DISCUSSION

(i) *Ambient medium discharges*

In all measurements of electric strength it is necessary to ensure that no discharges take place in the ambient medium of the electrode system, since such discharges generally lead to premature breakdown.

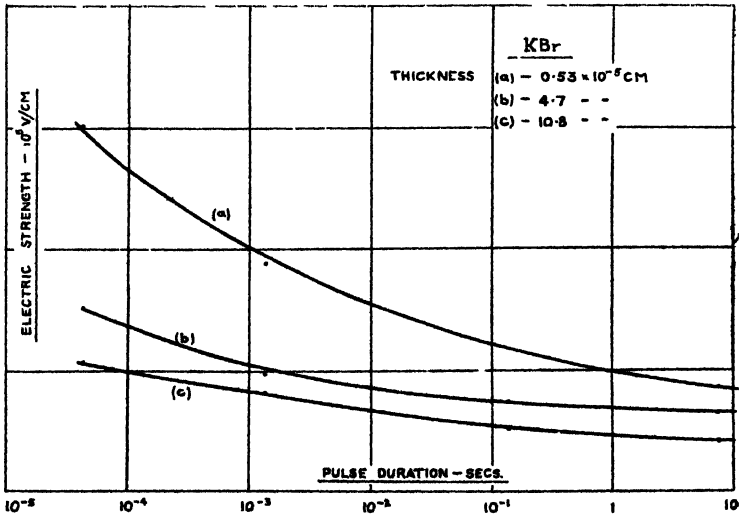


Figure 9.

For thick specimens air is unsuitable as an ambient medium, but Austen and Whitehead (1940) suggest that for thicknesses below about 6×10^{-5} cm. no adverse effects occur. The present measurements were carried out in air at a low pressure, but it was ascertained that the same results could be obtained in dry air at atmospheric pressure. The conditions of test were therefore similar to those of Austen and Whitehead, and their view is supported by a number of circumstances.

Thus, the electric strength values observed for nearly all the films are higher than those of the bulk material, measured under conditions which are claimed to preclude ambient discharges. Further, the effect of discharges would be expected to vary with film thickness and with pulse duration, whereas silica and polystyrene films show electric strengths independent of these two variables. Finally, the occurrence of discharges is generally accompanied by a large scatter in the readings of breakdown voltage, whereas a comparatively small scatter was observed with the crystalline films. From the theoretical point of view the absence of discharges in the case of thin films at atmospheric pressure is not obvious. In fact, on comparing the voltage V_A in the air gap between dielectric and electrode with the Paschen curve of spark-over voltage in air, V_S , for corresponding gap lengths, it is found that $V_A > V_S$ over an appreciable range of the gap, whenever the total voltage applied to the specimen is greater than the minimum value of 325 v. in the Paschen curve. Thus, discharges would be expected in such cases, in contradiction to the experimental evidence. The explanation of this disagreement may be sought in several effects:

- (a) The Paschen curve applies to spark-over in an irradiated gap between two metal electrodes. The gap in question being between a steel electrode and the dielectric may affect the spark-over potential and may raise the minimum.
- (b) The absence of irradiation may increase the apparent breakdown potential by introducing a time lag.
- (c) The nature of the discharge may be different for the very small gaps involved; for example, the pointed sparks, which are thought to be responsible for premature breakdown in the case of thick specimens, may give way to a relatively harmless glow discharge.

In the absence of data on discharges in small composite gaps, no conclusions can be drawn.

(ii) *The structure of films deposited by the vacuum evaporation process*

Limited information is available in the literature regarding the structure of deposited dielectric films, but the very considerable literature on metallic films affords a general picture which is useful as a basis for consideration of the probable structure in the dielectric case.

As a general rule, the films consist of an agglomerate of small crystals, whose lattice structure is the same as that of the bulk material. The extent to which the films are broken up depends on the size of the crystallites, and this is determined very largely by the nature and the temperature of the substrate on which the films are condensed, as well as by the particular metal of which the film is composed. The lower the substrate temperature, the smaller the crystal size, and vice versa.

Appleyard and Lovell (1937) by measuring the conductivity of very thin caesium films find evidence that at 64° K. the crystallization is inhibited altogether, but at higher temperatures crystallization takes place and the films break up into agglomerates.

The effect of both temperature and nature of the substrate is demonstrated by the behaviour of zinc films (Picard and Duffendack 1943), which are found to be reflecting when deposited on glass at liquid oxygen temperature, but of a dark bluish appearance when deposited at room temperature. Reflecting films can be obtained at room temperature by evaporating on copper or aluminium instead of glass or collodion. The interpretation of this effect is based on electron microscope studies of the films deposited on collodion. It appears that the reflecting films consist of crystals which are smaller than the wavelength of light and so form a relatively uniform surface, whereas the blue films consist of larger crystals which scatter the light instead of reflecting it.

The crystal size of some dielectric films has been estimated by Germer (1939) from the line width of electron diffraction patterns. Films of CaF_2 , NaF, CsI and CuCl were produced under conditions similar to those applying in the present work, but thin 'Formvar' films were used as substrate to allow electron diffraction in transmission. Because of the use of this substrate, which in addition may have been warmed up appreciably by the evaporating filament, it is thought that the films tested here may consist of crystals somewhat smaller than those investigated by Germer. For CaF_2 and NaF of average thickness 30–550 Å. Germer gives the crystal size as 100–150 Å., with a lower limit of 70 Å. CsI and CuCl have similar values, and presumably the crystal size of KBr will not be very different.

Taking the crystal size of the present films as 70 Å. approximately, it may be said that the thinnest CaF_2 and KBr films tested ($t=500$ Å.) should present a reasonably uniform surface, while the thinnest NaF film ($t=230$ Å.) may be seriously broken up into agglomerates.

Evaluation of the refractive index of the various films from the interferometric data referred to in § 3 (iii) gives values somewhat lower than for the corresponding single crystals and indicates that the films consist of aggregates of tiny crystals with voids between them. Thus, for CaF_2 , 1.29–1.35 was found instead of 1.41, NaF gave 1.29–1.31 instead of 1.32, and only KBr gave values close to the true refractive index of 1.56.

The effect of this film structure on the mechanism of breakdown may be two-fold. If the electrons suffer additional collisions at crystal boundaries, then the mean free path, which is rather larger than the crystal size, will be reduced, and the electric strength should be increased. No such effect will take place if the bridges between crystallites are so perfect that no additional scattering occurs.

From a more macroscopic point of view, the porosity may give rise to local field concentrations and so cause premature breakdown.

The structure of silica films is not known, but, if the rule that evaporated films have the same lattice structure as the bulk material may be generalized, then the silica films should be amorphous.

(iii) *The structure of polystyrene films*

Polystyrene is a non-polar hydrocarbon polymer, the chain molecules of which are about 1000 carbon atoms long; it is in this respect rather similar to

polythene, for which some work on the structure of thin films has been published. Charlesby (1945) produced thin films of polythene (approximately 1000 Å. thick) by placing a drop of a dilute solution on water at 90°C. and allowing the solvent to evaporate. From the electron diffraction pattern obtained in transmission, and from consideration of the shape of the polythene molecule it was deduced that the molecules were oriented so as to make an angle of 56° with the plane of the film, and were tightly packed forming small crystals. When preparing the film by placing a drop of solution on a metal surface, less orientation resulted.

Orientation in polystyrene may be further inhibited in view of the bulky side chains on the molecules compared with polythene, and so the films may be mainly amorphous. The effect of any orientation of the chains into the direction of the field would, according to Fröhlich, be to lower the electric strength.

(iv) *The experimental results in relation to theory*

The experimental results fall into two groups, those on crystalline films, where a variation of electric strength both with pulse time and thickness was found, and those on silica and polystyrene films for which no such variation was observed.

Considering the latter group first, the absence of a variation of electric strength over the thickness range tested is in agreement with Fröhlich's theory, if the films are assumed amorphous in structure. This is probably the case. The average value of 5.5×10^6 v/cm. found for polystyrene compares well with the electric strength of 6.1×10^6 v/cm., determined by W. G. Oakes* for recessed specimens of thickness about 0.04 mm. The lowering by 10% is barely significant, but may be due to orientation of the molecules. The average electric strength found for silica films, 12.5×10^6 v/cm., is appreciably higher than the values found by von Hippel and Maurer (1941) for a number of thick specimens, their highest figure being 8×10^6 v/cm. for a thickness of 6×10^{-4} cm.

Several explanations are possible:—

- (a) Silica has been observed to give varying results depending on the particular specimen and its history (von Hippel and Maurer 1941), and the difference between the figures quoted may be due to this cause.
- (b) In spite of the precautions taken, some tungsten from the filament may have evaporated simultaneously with the silica and introduced metallic centres into the film. According to Fröhlich (1937, 1939, 1941), this should reduce the mean free path and raise the electric strength.
- (c) The scatter accompanying the results does not allow a definite line to be drawn, and the figure of 8×10^6 v/cm. at 6×10^{-4} cm. thickness may, in fact, lie on the continuation of a curve which is inclined rather more than that drawn in figure 6.

The case of the crystalline materials CaF_2 , NaF and KBr is complicated by the time dependence of the electric strength, which is not predicted by Fröhlich's theory. At a given value of the pulse duration, however, the electric strength rises with decreasing thickness, as shown in figures 7, 8 and 9, and in columns (2) and (3) of table 3. The electric strengths of the thickest films at the longest

* Data kindly communicated to the author prior to publication.

pulse durations agree well with the values found by von Hippel (1937) for thick specimens. No exact calculation of the rise in electric strength to be expected for thin films has been carried out by Fröhlich and therefore no direct correlation with theory is possible. On the other hand, a numerical comparison based on the calculated mean free paths appears desirable and the final column of table 3 headed 'Estimated value' has been drawn up, assuming Austen's result on mica to apply, viz. a 60% increase in strength at a thickness of 40 mean free paths. Thus, this column gives the observed limiting electric strengths at large thickness and long pulse durations increased by 60%, which may then be compared with the corresponding experimental figures at a thickness of 40 mean free paths shown in the two preceding columns for two different pulse durations. It is seen that the estimated values lie between the experimental figures for the two pulse durations.

Table 3

	Thickness	Electric strength values in 10^6 v./cm.			
		(1) von Hippel's value	(2) Experimental value Pulse duration		(4) Estimated value
			10 sec.	10^{-4} sec.	
CaF ₂	$6 \cdot 10^{-5}$ cm. (40 <i>l</i>)	—	2.8	4.6	4.5
	10^{-5} cm.		2.8	6.4	—
NaF	4.4×10^{-5} cm. (40 <i>l</i>)	} 2.4	2.6	5.7	3.9
	10^{-5} cm.		3.2	7.1	—
KBr	10^{-4} cm. (40 <i>l</i>)	} 0.75	0.75	2.1	1.2
	10^{-5} cm.		1.6	4.8	—

l = mean free path.

(v) *Possible explanations of the time effect*

There are two major possibilities :—

- (a) the effect is inherent in the properties of thin crystalline layers,
- (b) the effect is due to some undesired influence.

Although not brought out explicitly, figures 7, 8 and 9 show that there is a tendency for the time effect to disappear as the film thickness increases. The reality of the effect is therefore possible, since it would not be noticed with thick specimens. At the same time, an explanation on the basis of electronic breakdown is difficult to find, since one would expect electronic movements to be completed in times of the order of 10^{-8} sec. or less. Possibly a connection could be sought with the shortness of the path available for the acceleration of the electrons

and the consequent necessity for a number of relatively improbable collisions to occur before instability sets in.

Any possibility that the time dependence might be associated with some characteristic of the testing circuit, a crushing effect on the test specimen arising from electrostatic attraction between the electrodes, electrolytic decomposition of the film and ambient medium discharge has been ruled out either by test or by rough calculation. There remain the possibilities of impurities in the films causing increased conductivity and heating, or some interfacial polarization effects due to the polycrystalline structure of the films. An attempt was made to probe into the question of impurities by evaporating a different sample of CaF_2 from a tungsten instead of a molybdenum heater. Although the results agreed with the previous ones within experimental error, this does not constitute proof of the absence of impurities. Conductivity measurements and tests with shorter pulses, when at some stage the heating effect should become negligible, would be better tools to employ.

§ 6. CONCLUSIONS

It may be said that the results obtained agree qualitatively with Fröhlich's theory, which predicts a rise in electric strength as the specimen thickness approaches the electronic mean free path.

In view of the uncertainty whether evaporated films with their polycrystalline structure are truly representative of the ideal lattice postulated in theory, and because of the time effect, the cause of which is not yet known, it is not possible to draw any definite conclusions.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor Willis Jackson for constant guidance and encouragement and to Dr. W. K. Donaldson for help with the optical work. Thanks are due to the Department of Scientific and Industrial Research for financial assistance in the form of a special research grant.

REFERENCES

- APPLEYARD, E. T. S., and LOVELL, A. C. B., 1937, *Proc. Roy. Soc. A*, **158**, 718.
 AUSTEN, A. E. W., and HACKETT, W., 1939, *Nature, Lond.*, **143**, 637.
 AUSTEN, A. E. W., and WHITEHEAD, S., 1940, *Proc. Roy. Soc. A*, **176**, 33.
 CHARLESBY, A., 1945, *Proc. Phys. Soc.*, **57**, 496.
 FRÖHLICH, H., 1937, *Proc. Roy. Soc. A*, **160**, 230; 1939, *Ibid.*, **172**, 94; 1941, *Ibid.*, **178**, 493.
 GERMER, L. H., 1939, *Phys. Rev.*, **56**, 58.
 HILSCH, R., and POHL, R. W., 1930, *Z. Phys.*, **59**, 812.
 VON HIPPEL, A., 1937, *J. Appl. Phys.*, **8**, 815.
 VON HIPPEL, A., and MAURER, R. J., 1941, *Phys. Rev.*, **59**, 820.
 PICARD, R. G., and DUFFENDACK, O. S., 1943, *J. Appl. Phys.*, **14**, 291.
 PLESSNER, K. W., 1946, *Nature, Lond.*, **158**, 909.
 TOLANSKY, S., 1945, *J. Sci. Instrum.*, **22**, 161.
 TURNER, C. H. M., and LEWIS, W. E., 1947, *Nature, Lond.*, **159**, 334.

Excited Electronic Levels in Conjugated Molecules: I. Long Wavelength Ultra-Violet Absorption of Naphthalene, Anthracene and Homologs

By C. A. COULSON

Physical Chemistry Laboratory, Oxford*

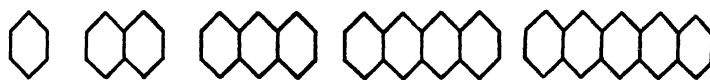
* Now at King's College, London

MS. received 28 June 1947

ABSTRACT. The method of molecular orbitals has been applied to study the long wave ultra-violet absorption of the polyacene series, naphthalene, anthracene, The longest wavelength in an allowed excitation probably arises from a transition between mobile electron levels of symmetry $A_{1g} \rightarrow B_{2u}$, and polarized across the width of the molecule. In the first four or five members of the series, the second wavelength is due to a transition $A_{1g} \rightarrow B_{1u}$, polarized along the length of the molecule. Discrepancies between this theory and both the valence-bond resonance treatment and the Lewis-Calvin theory of colour are elucidated. This assignment of transitions is in agreement with (1) calculation of absolute energies as a function of chain length, (2) calculation of relative energies of the B_{2u} and B_{1u} transitions, (3) limiting condition for a large number of rings, (4) calculations of intensities, and (5) some experimental work on fluorescence.

§ 1. INTRODUCTION

THE series of molecules shown in figure 1 may be called the polyacene series, individual members being defined in terms of the number n of rings. This means that the molecule n has a formula $C_{4n+2}H_{2n+4}$. Their long wavelength u.v. absorption spectra have been measured from benzene with one



benzene naphthalene anthracene naphthalene pentacene
Figure 1.

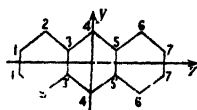


Figure 2.
Notation in anthracene.

ring ($n=1$) up to pentacene with five rings ($n=5$). The close similarity of these spectra (see e.g. Bowen, 1946, p. 127, where they are reproduced on the same diagram) shows that a similar process is responsible for them all. There is, indeed, a characteristic shift to longer wavelength with increasing size of molecule. Thus each additional ring gives rise to a red shift of about 1000 Å. from naphthalene (2750 Å.) to pentacene (5800 Å.). This shift proves that the process must be the excitation of a conjugation, or mobile, electron, so that the absorption is of the kind called $N \rightarrow V$ by Mulliken (Mulliken and Rieke 1941, subsequently referred to simply as M.R.). Our object in this paper is to show that the magnitude of the red shift can be calculated by the theory of molecular orbitals; and we shall be led to quite definite conclusions regarding the direction of the transition moment.

This, is of course, simply the direction of the electric vector in the absorbed, or fluorescent, light. Since the transition is $N \rightarrow V$, the electric vector must lie in the plane of the molecule. This conclusion has been explicitly verified for the related molecule chrysene (Krishnan and Seshan 1939), for anthracene (Krishnan and Seshan 1936) and for naphthacene (Scheibe 1938, Scheibe and Kandler 1938). We shall have occasion to refer to some of this work later. Now the fact that the transition moment lies in the plane of the molecule does not complete our knowledge. For if, as in figure 2, we take the x -direction perpendicular to the plane,* the transition moment may lie along the z or y directions. In this it differs from a straight polyene chain, or a linear dye molecule, where the moment must lie along the chain length. The classical analogy is to electrons oscillating in the y - or z -directions. It has hitherto been customary (e.g. Förster 1938) to suppose, by analogy with the conjugated polyenes, that the moment is along the greater axis of the molecule, which we have called z ; and, indeed, the valence-bond treatment of Förster, in so far as it is valid, supports this view, as also, for the special case of naphthalene, do some recent improved calculations of Mr. D. Craig, now in course of publication. But we shall show that the molecular orbital treatment leads to a transition moment in the y -direction. Here then is an interesting case where the two main methods disagree in a specific prediction. For that reason we have set out all the arguments that we can find, so that all possible evidence may be used to form a final judgment. It does not seem possible to decide unequivocally one way or the other, though we shall show that the molecular orbital (m.o.) calculations are supported by no less than five distinct lines of argument. It would seem, therefore, that they have a considerable plausibility. These arguments are (i) calculation of absorption energies as a function of chain length, (ii) calculation of relative energies for the $N - V_1$ and $N - V_2$ transitions, (iii) limiting condition for large molecules, (iv) calculation of intensities, and (v) some experiments on fluorescence. But the matter is not finally and completely settled.

§ 2. SYMMETRY PROPERTIES

In the notation of Mulliken (1933) all the molecules of this series are plane and have symmetry $V_h(-D_{2h})$. This means that there are only eight distinct types of m.o., shown in the character table below as A_{1g}, \dots, B_{3u} . This table, devised in the manner of Mulliken, shows how each of the eight types behaves under the influence of the eight sets of covering operations of the symmetry group. It also shows the behaviour of the coordinates x, y, z . It is not difficult to show that provided we deal entirely with mobile electrons (that is, with electrons whose wave function is antisymmetrical with respect to reflection in the plane of the molecule, and may therefore be built up, approximately, from $2p_z$ atomic orbitals), then for the individual electrons, though not for the totality of mobile electrons, we need only consider four of the eight types given in the table. These are $A_{1u}, B_{1g}, B_{2g}, B_{3u}$. Now in a molecule such as pentacene there are 22 mobile electrons, occupying several orbits of each of the four types. Our notation will subsequently be much simplified if, instead of using the full names A_{1u}, \dots , we use other symbols P, Q, R, S for each separate electronic m.o., reserving the

* This convention for the x axis is by no means invariable. Sometimes the directions of the x and z axes are interchanged from figure 2; and then B_{1g}, B_{1u} etc. become B_{3g}, B_{3u} and *vice versa*; but A_1 and B_2 are unchanged.

fuller description for the molecular states that result from the totality of the mobile electrons. The correlation between the two sets of names is shown below.

Table 1. Character table for $V_h \equiv D_{2h}$

	Coordinates	E	$C_2(z)$	$C_2(y)$	$C_2(x)$	i	$iC_2(z)$	$iC_2(y)$	$iC_2(x)$
A_{1g}		1	1	1	1	1	1	1	1
A_{1u}		1	1	1	1	1	-1	1	-1
B_{1g}		1	1	-1	-1	1	1	-1	-1
B_{1u}	z	1	1	-1	-1	-1	-1	1	1
B_{2g}		1	1	1	1	1	-1	1	-1
B_{2u}	y	1	1	1	-1	-1	1	-1	1
B_{3g}		1	1	-1	1	1	-1	-1	1
B_{3u}	x	1	-1	-1	1	-1	1	1	-1

Simplified notation P Q R S
 Full group-theory notation . . . B_{3u} B_{2g} B_{1g} A_{1u}

One advantage of the distinction that we are making is that it will always be obvious whether we are speaking of a single electron, or of the whole molecule. Another is that we can distinguish the m.o. of a given type (e.g. P) by a suffix, showing its order on an energy scale. Thus P_2 would denote the second m.o. of type P , and the long wavelength absorption in anthracene (see §3) would be written

$$P_1^2 Q_1^2 R_1^2 P_2^2 Q_2^2 S_1^2 R_2^2 \quad {}^1A_{1g} \rightarrow P_1^2 Q_1^2 R_1^2 P_2^2 Q_2^2 S_1^2 R_2 P_3 \quad {}^1B_{2u}. \quad (1)$$

In this description the electrons of greatest binding energy come first (though in this particular case R_1 and P_2 have equal energy, and so do Q_2 and S_1 . This, as figure 3 shows, is unusual.) Equation (1) shows that absorption corresponds essentially to the excitation of an electron in orbit R_2 to a higher orbit P_3 . A convenient abbreviation would be

$$R_2^2 \quad {}^1A_{1g} \rightarrow R_2 P_3 \quad {}^1B_{2u}. \quad \dots\dots (2)$$

Since, as table 1 shows, two electrons in any orbit have combined symmetry A_{1g} , the selection rules and symmetry classes for the absorption jump are in fact entirely determined by the orbits shown in (2). Now, in what follows, we shall only be concerned with molecules in which, on excitation, there are two singly-occupied m.o. This means, in effect, that only one electron is excited from the ground state configuration. Table 2 shows the symmetry representation corresponding to any

Table 2. Symmetry class of product of two electrons

		Q	R	
P	A_{1g}	B_{1u}	B_{2u}	B_{3g}
Q	B_{1u}	A_{1g}	B_{3g}	B_{2u}
R	B_{2u}	B_{3g}	A_{1g}	B_{1u}
S	B_{3g}	B_{2u}	B_{1u}	A_{1g}

two electrons in any pair of orbitals of the types P , Q , R or S . Thus the intersection of the row R and the column P gives the type (B_{2u}) associated with the

unpaired electron configuration RP , as in (2). There is, of course, usually both a singlet and a triplet level, e.g. $^1B_{2u}$ and $^3B_{2u}$; but we shall only be concerned with the singlet, since the ground state is always a singlet $^1A_{1g}$, and singlet-triplet transitions are not allowed. Now the ground state is A_{1g} , and allowed transitions must satisfy the $g \longleftrightarrow u$ selection rule. Table 2 shows that this restricts excitation except to B_{1u} and B_{2u} configurations. The allowed upper states and the corresponding directions of polarization are shown in table 3.

Table 3. Allowed upper states and direction of polarization

Upper state	Absorption process	Polarization
PR , or QS	$A_{1g} \rightarrow B_{2u}$	y
PQ , or RS	$A_{1g} \rightarrow B_{1u}$	z

With this background, which is common to all the molecules of this series, we may now pass to a discussion first of the energy change in absorption, then of the direction and intensity of the transition. It will suffice to explain the analysis for one molecule, anthracene, in detail. The others are entirely similar.

§ 3. ANTHRACENE

Let us number the carbon atoms of anthracene as shown in figure 2. This is not the conventional chemical numbering, but it is the one best suited to this problem. Also, let ψ_j denote the $2p_x$ orbital of carbon atom j . Then we are to form the various allowed m.o. by linear combination of all the ψ_j and $\psi_{j'}$. Alternatively, as was shown by Hückel (1932), we may choose any suitable groupings of these. Now, by symmetry, $\psi_1, \psi_{1'}, \psi_7, \psi_{7'}$ must always occur with equal weight, and similarly for the others. This leads us to the following sub-groups, in which we have omitted the ψ and used simply j to stand for ψ_j . These are seen to conform to the symmetry classes discussed in § 2.

$$\left. \begin{array}{l} \text{Group } P, \quad 1+1'+7+7', \quad 2+2'+6+6', \quad 3+3'+5+5', \quad 4+4', \\ \text{Group } Q, \quad 1+1'-7-7', \quad 2+2'-6-6', \quad 3+3'-5-5', \quad . \\ \text{Group } R, \quad 1-1'+7-7', \quad 2-2'+6-6', \quad 3-3'+5-5', \quad 4-4', \\ \text{Group } S, \quad 1-1'-7+7', \quad 2-2'-6+6', \quad 3-3'-5+5'. \end{array} \right\} \dots\dots (3)$$

On account of their symmetry, the elements of different rows are non-combining, so that the allowed m.o. are formed solely out of members of any one row. Thus there are four m.o. found by suitable linear combination of the four members in the top row; all these are of P type and may be called P_1, P_2, P_3, P_4 . The energies may be found by finding the roots of the appropriate secular determinants, as was done by Hückel (see also Wheland 1941). In the simplest approximation, where overlap between adjacent atomic orbitals is neglected, these energies are of the form

$$E = E_0 + k\beta, \quad \dots\dots (4)$$

where E_0 is the energy of a $2p_x$ electron on any one nucleus, β is a resonance integral between adjacent nuclei, and k is a numerical coefficient different for each level and found from the secular determinant. As β is numerically negative, positive values of k correspond to bonding, negative values to anti-bonding, orbitals.

In the complete set of levels (3), positive and negative values of k occur equally often. If we do not neglect the overlap S between adjacent atomic orbitals, the energies are no longer given by (4), but, as Wheland showed, they are

$$E = E_0 + m\gamma, \quad \text{where } m = k/(1 + kS). \quad \dots\dots(5)$$

Here γ is a new resonance integral which replaces β , and is equal to $E_0 - S\beta$. In calculations, S is given the value 0.25. Our procedure, therefore, is to solve the simpler equation to get the values of k in (4); then we use (5) to convert from k -values to m -values. In this way the energies shown in figure 3 are obtained.

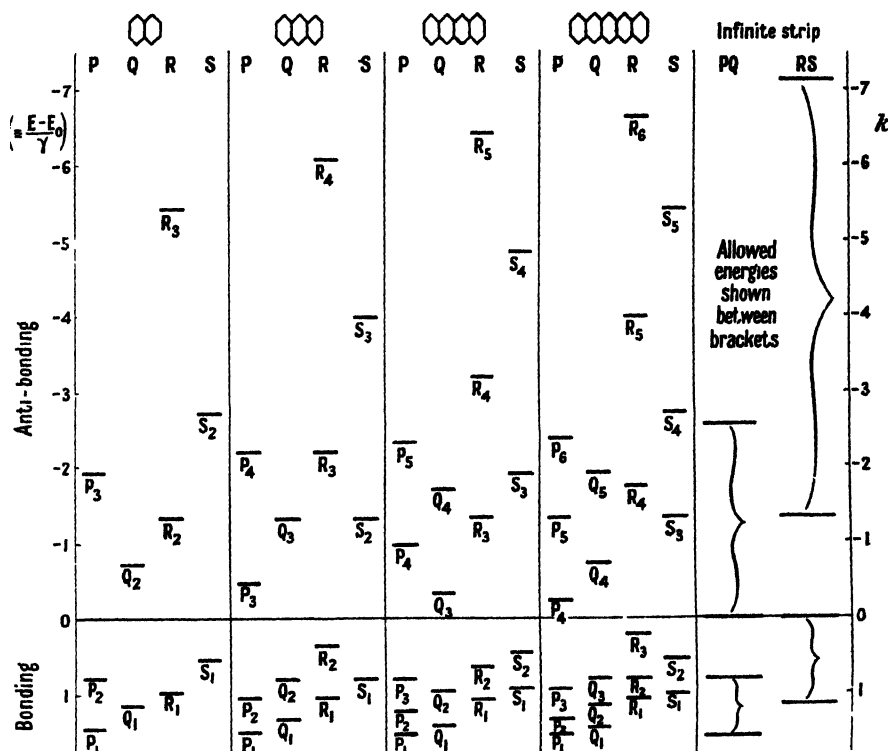


Figure 3.

This shows the allowed energies in each of the four groups P - S , plotted in terms of γ . The solid horizontal line corresponds to $E = E_0$, i.e. $m = 0$. Orbitals below the line are bonding, those above it are anti-bonding. Notice the very considerable asymmetry in the two halves—a situation already noted in the case of the continuous band distributions of infinite strips of carbon atoms by Coulson and Rushbrooke (1947). The simpler molecules of this series give secular equations which are fairly easily solved directly, but for the more complex members, and for the infinite polyacene chain (see later), the method explained in the appendix is useful.

In the ground state, all seven bonding orbitals are doubly occupied, leading to the assignment stated in (1).

Excited states are reached by taking one of the electrons from an orbital below the line and placing it in an orbital above the line. The excitation energy is

obviously least if the transition is $R_2 \rightarrow P_3$: or, more accurately, $R_2^2 \rightarrow R_2P_3$, as stated in (2). A glance at table 3 shows that this is an allowed transition with polarization along the y -direction. There are, of course, many other possible allowed transitions. Following M.R.'s notation we may call them $N \rightarrow V_1$, $N \rightarrow V_2$, The five lowest, including two forbidden ones, are shown below.

Table 4. $N \rightarrow V$ transitions in anthracene

Transition	Description	Excited configuration	Excitation energy	Polarization
$N \rightarrow V_1$	$R_2^2 \rightarrow R_2P_3$	B_{2u}	0.836 γ	y
$N \rightarrow V_2$	$Q_2^2 \rightarrow Q_2P_3$	B_{1u}	1.261 γ	z
$N \rightarrow V_3$	$S_1^2 \rightarrow S_1P_3$	B_{3u}	1.261 γ	forbidden
$N \rightarrow V_4$	$R_1^2 \rightarrow R_1P_3$	B_{2u}	1.506 γ	y
$N \rightarrow V_5$	$P_2^2 \rightarrow P_2P_3$	A_{1g}	1.506 γ	forbidden

Interaction between the lowest B_{2u} states V_1 and V_4 will slightly lower the $N \rightarrow V_1$ energy. So also will interaction between V_2 and V_6 and V_7 (all type B_{1u}) lower V_2 . It is not the purpose of this paper to calculate these additional energy changes, which are zero in our approximation, and small in any case. Certainly they could not be responsible for driving the B_{1u} level below the B_{2u} . The position here is quite clear-cut: the y -polarization is of lowest energy.

§4. OTHER MOLECULES

Figure 3 shows the energy values for the other molecules of this series with $n=2, 4, 5$. They are obtained in just the same way and are grouped into their P, Q, R, S types. In the molecules with odd values of n (such as anthracene) the lowest absorption is $R^2 \rightarrow RP$; in the even molecules it is $S^2 \rightarrow SQ$. Now both of these are allowed transitions, of the form ${}^1A_{1g} \rightarrow {}^1B_{2u}$, with polarization in the y -direction. As naphthalene is more important than the others, we give, below, the lowest five transitions for this molecule, similar to table 4 for anthracene.

Table 5. $N \rightarrow V$ transitions in naphthalene

Transition	Description	Excited configuration	Excitation energy	Polarization
$N \rightarrow V_1$	$S_1^2 \rightarrow S_1Q_2$	B_{2u}	1.267 γ	y
$N \rightarrow V_2$	$P_2^2 \rightarrow P_2Q_2$	B_{1u}	1.532 γ	z
$N \rightarrow V_3$	$R_1^2 \rightarrow R_1Q_2$	B_{3u}	1.715 γ	forbidden
$N \rightarrow V_4$	$S_1^2 \rightarrow S_1R_2$	B_{2u}	1.868 γ	z
$N \rightarrow V_5$	$Q_1^2 \rightarrow Q_1Q_2$	A_{1g}	1.884 γ	forbidden

Such tables as 4 and 5 are easily written down for other members of the series. They all show a y -polarization as having the lowest energy, the closest z -polarization being at an increasing distance above the y -polarization as n increases. Figure 4 shows how the excitation energies of the B_{2u} and B_{1u} transitions vary from molecule to molecule. In benzene, with its extremely high symmetry the simple division valid for the other molecules no longer applies. But th

corresponding point at 2.133γ may be said to belong to both groups, and we have drawn the two curves accordingly. At large values of n , it may be shown (see Appendix) that the lower (i.e. B_{2u}) curve tends to zero energy, whereas the upper (i.e. B_{1u}) curve tends to a finite value 0.8γ . We have, in fact, a case of partially overlapping bands, as found in the theory of metals (see also Coulson and Rushbrooke 1947). In the case of a large molecule we are effectively dealing with a crystal strip, and our method

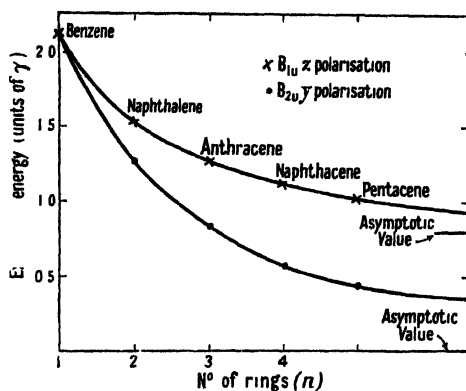


Figure 4.

is synonymous with that known as the Bloch theory. This has given such good qualitative results in the theory of metals that it is almost inconceivable that its prediction for the large strip should be sufficiently inaccurate for the B_{1u} level to lie below the B_{2u} one. We may, indeed, regard this as substantial evidence in favour of the assignment of direction that we have made for the long wavelength transition.

This conclusion concerning the direction of polarization of absorbed light is unexpected. But it receives confirmation in some experimental work of Krishnan and Seshan (1936). These workers studied the fluorescence of small traces of naphthacene imbedded in solid solutions of anthracene or of chrysene. They found (i) that the absorption was much stronger when the direction of electric vibration was in the plane of the rings, (ii) that "whether the exciting light vibrations are along the b axis, or along a, the direction of polarization of fluorescent light remains the same, the component of the fluorescent light vibrating along the b axis being much more intense than the component along the a axis"; and (iii) "the b vibration is predominant in the fluorescent light not only when the exciting light vector is along the same direction—a result normally to be expected—but also when the light vector is along a, which is perpendicular to b—a very striking result". They called this phenomenon negative polarization. Now the molecules in the crystal (see e.g. Robertson 1933) are unfortunately not all parallel, but in naphthacene and anthracene the molecules are nearly normal to a, so that the b direction is quite close to the short direction in the plane—the direction which we have been calling γ . This shows almost conclusively, first, that the upper state of the long wave absorption is a B_{2u} state; second, that absorption along the a axis, which may lead either to x or z polarization, requires greater energy; and third, that after excitation, some of this greater energy may be degraded with final emission of γ -polarized light. All this is in complete agreement with our theoretical calculations.

But there is a further test that we may apply. We have seen that in every case the m.o. theory predicts that the V_1 state is B_{2u} , and (at least for the smaller molecules) the V_2 state is B_{1u} . It is possible (Jones 1945, Clar 1933) to observe the approximate positions of the first two u.v. bands and we may compare the ratio of the energies $(N \rightarrow V_2)/(N \rightarrow V_1)$ for the series of molecules naphthalene to pentacene. Table 6 shows both the observed and calculated values for this ratio. The agreement is good, judged on an absolute scale. But when it is recognized

that the variations of this ratio with the number of rings is not what would have been expected *a priori*, and yet theory and experiment are nowhere more than 15% apart, it must be admitted that there is very strong support for the interpretation of these transitions that we have given. In view of this it is hard to avoid the

Table 6. Variation of energy ratio $(N \rightarrow V_2)/(N \rightarrow V_1)$ with number of rings

Molecule	No. of rings	Experimental ratio	Calculated ratio
Naphthalene	2	1.14	1.21
Anthracene	3	1.48	1.51
Naphthacene	4	1.70	1.89
Pentacene	5	2.06	2.34

conclusion that the B_{2n} transition is the lowest, and that as n increases, λ increases without limit, giving what is known as a divergent sequence. But the energy of the B_{1n} transition tends to a finite non-zero value; in other words, this sequence is convergent. The question now arises—how accurate are the absolute values of the calculated absorption frequencies? There are several reasons why a great accuracy is not to be expected. First, the calculated frequency is directly proportional to the resonance integral γ , and the value of this quantity is not known with complete reliability. Second, the bond lengths, and hence the individual values of γ , have all been assumed equal, whereas Moffitt and Coulson in a paper in course of publication have shown that there are fairly large deviations from equality. Third, and most serious, we have completely neglected the singlet-triplet separation in the excited state. The writer will show in a later paper in this series that for polyene chains the excited singlet lies considerably above the triplet, the average of the two being given by the treatment we have here been using. It seems reasonable to suppose that the same relative separations are likely to occur for polyacenes as for polyenes. If this is so, corrections of up to 20%, or even 30%, have to be introduced. But if we are solely concerned with one set of level either the singlets or the triplets, but not both together, these corrections may be almost completely obviated by a simple change in the numerical value of the resonance integral γ ; as a result we may, with some accuracy, calculate the transition energies by neglecting the spin interaction which leads to a distinction between singlet and triplet, by the simple expedient of using a different γ for the singlets and for the triplets. As we are here only interested in the singlets, this amounts to saying that the value of γ required for these u.v. transitions in our simple theory is of the order of 20 to 30% greater than the value of γ required to compare resonance energies and heats of formation. When this is admitted, we find a quite reasonable agreement. For example, the correct wavelength (2750 Å.) for naphthalene requires γ to be about 3.5 e.v. If we subtract 30% we get 2.3 e.v., a value lying within the range 2 to 2.5 e.v. commonly used for this integral in resonance calculations.

If the above argument is correct, we should be able to predict with considerable more reliability the characteristic red shift with increasing length of molecule since we may express our results in a form independent of the precise numerical value of γ . This is indeed the case. Table 7 shows how the ratio of the w

lengths of the $N \rightarrow V_1$ transition in a given molecule (n) of this class and in naphthalene ($n=2$) varies with n . The agreement between theory and experiment is as satisfactory as we should expect from this kind of treatment.

Table 7. Ratio of lowest absorption wavelength to that of naphthalene

Calculated ratio	1.00	1.51	2.11	2.86
Observed ratio	1.00	1.35	1.67	2.11

It may be mentioned that for large values of n , this theory predicts a variation of the absorption wavelength of the form

$$\lambda = \frac{h}{2\pi^2\nu}(n+1)^2.$$

§ 5. INTENSITIES

A further test of the theory is found by calculating the intensities of the various transitions. This (see M.R. pp. 232-3) is achieved by means of either the dipole strength D , the transition moment Q or the oscillator strength f . Table 8 shows the values obtained in this way. The column headed f_{theor} represents the oscillator strengths deduced from the equation (M.R. equation (2))

$$f = 1.085 \times 10^{11} \times \nu \times D,$$

and using the experimental values of ν shown in the previous column, which are quoted from Förster (1938). The column headed f_{exp} represents values deduced from experimental extinction curves. The writer gratefully acknowledges the provision of these, prior to publication, by Mr. D. Craig.

Table 8. Intensities of B_{2u} transition

Molecule	n	$Q(\text{A.})$	$D(\text{A.}^2)$	$\nu(\text{cm.}^{-1})$	f_{theor}	f_{exp}	$D(B_{1u})$
Naphthalene	2	1.706	2.910	36 400	1.15	0.11	2.167
Anthracene	3	1.755	3.080	26 600	0.89	0.09	3.316
Naphthacene	4	1.796	3.226	21 700	0.76	0.08	4.331
Pentacene	5	1.829	3.345	17 200	0.63	0.08	5.185

We may notice several features in table 8. First, the dipole strength is practically constant along the series. This is to be expected if the transition is along the short length of the molecule, but not if it is along the greater length. Verification of this is found in the last column, where we show the dipole strength calculated for the lowest $A_{1g} \rightarrow B_{1u}$ transition. Here D increases almost regularly with n , a situation found in almost identical terms by M.R. (table 5 on p. 258) for the diphenyl polyenes $\text{Ph}-(\text{CH}=\text{CH})_n-\text{Ph}$, where the transition moment must necessarily lie along the length of the chain. This almost steady value of D in the B_{2u} transition leads to a slight decrease in the value of f as n increases. This decrease is just the same as that found experimentally and provides further justification for the B_{2u} interpretation. It is perhaps worth mentioning that a

consideration of the coefficients in the relevant m.o. shows that if all the C-C bonds are supposed to have the same length r_{C-C} (in Å.), then the $N \rightarrow I^+$ transition moments for all $A_{1g} \rightarrow B_{2u}$ transitions are such that Q lies between $\sqrt{2}r_{C-C}$ and $2\sqrt{2}r_{C-C}$. There is no such limitation in the $A_{1g} \rightarrow B_{1u}$ transitions, for which Q appears to increase almost linearly with \sqrt{n} .

The absolute value of f , as calculated, is much too high. This is a familiar situation. Thus M.R. (p. 263) introduce correction factors varying from 2/5 to 1/4, and for F_2 and Cl_2 (p. 251) a correction factor of 0.1 is required. This is of the same order as our ratio between f_{exp} and f_{theor} . It is a pity that such accuracy is about the best that can be expected if we are not prepared to consider more carefully the value of the overlap integrals, changes in bond length and effect of vibrations. The significant fact about the column f_{theor} is that the ratio $f_{\text{theor}}/f_{\text{exp}}$ is substantially constant for the whole series of molecules.

A word may now be said about the relation of this work to the semi-classical theory of colour developed by Lewis and Calvin (1939). These authors speak of electron oscillations through the whole, or a part, of the system, and refer to absorption bands of complete or of partial oscillation. In the case of polyacenes, bands of complete oscillation would be those for which, in quantum language, the transition moment lay along the z -axis, and corresponded to a "charge transfer" across a fair proportion of the total length. The theory developed by Lewis and Calvin postulates that the restoring force on an electron moving along the greater length is less than along the shorter, so that the oscillations of least energy are in this direction; and therefore the long wavelength spectrum should be polarized this way. Our theory runs entirely counter to this view. For however long the molecule may be, the charge transfer is along the short direction in the molecular plane: and the observed red shifts with increasing size are accounted for quite adequately without recourse to the semi-quantum picture of Lewis and Calvin. Indeed, this latter theory cannot possibly explain what we believe to be the case, namely that the red shift is associated with oscillations *along* the chain length in polyenes and *across* the chain length in polyacenes (cf. particularly Lewis and Calvin, pp. 293-4, 314). It seems as if this theory must be abandoned. We can, in fact, see fairly easily where it goes wrong. Quantum-mechanically we know that there are various allowed energies for the molecular electrons; and the lowest energy is one in which (see e.g. Bowen 1946, pp. 106-113) there is no node in the wave function. Increasing energies are associated with increasing numbers of nodes. These nodes, like the nodes in a vibrating rectangular plate, are parallel to one or other of the axes of the molecule. In the ground state of the molecule all the lowest levels are filled, so that an electron making a transition of the kind we are discussing must already have a wave function with several nodes. Whether the extra node that it acquires on excitation is in the one direction or the other (i.e. whether the polarization is in the one direction or the other) cannot possibly be deduced from the kind of argument used by Lewis and Calvin, unless it be supposed, on some kind of parallel with a harmonic oscillator, that the energy difference between successive levels of the same kind (partial or complete) is effectively unchanged for the first seven or eight levels of each kind. Such a hypothesis is manifestly wrong, as the whole body of spectral evidence from atoms shows. All that these authors do is to determine whether it is easier, starting from a nodeless wave function, to create a node in the one direction or the other. They conclude that it is easiest to create a node which divides the long axis of the molecule in two.

This will be recognized, in figure 3, as a statement that when we start in the level P_1 (which has no nodes of the kind we are discussing) it requires less energy to go to Q_1 than to R_1 ; and, what is more, the energy separation P_1 to Q_1 decreases as $1/n$. Lewis and Calvin's error is that they have confused the level of the lowest possible molecular orbital with that of the highest occupied molecular orbital. For that reason their theory is quite invalid in this particular type of molecule.

§ 6. COMPARISON WITH VALENCE-BOND TREATMENT

We close with an account of the relation between our molecular orbital treatment and the corresponding valence-bond treatment, the only specific published calculations for which are due to Förster (1938). Let us consider as a particular example, the anthracene molecule. Förster regarded the ground state and the lowest excited state as combinations of the four unexcited (Kekulé) structures $A \dots D$. Their wave functions ψ_I and ψ_{II} were of the forms

$$\psi_I = c_1(\psi_A + \psi_B) + c_2(\psi_C + \psi_D),$$

$$\psi_{II} = c_3(\psi_A - \psi_B) + c_4(\psi_C - \psi_D),$$

with $c_1 \dots c_4$ as determined numerical constants. The transition $\psi_I \rightarrow \psi_{II}$

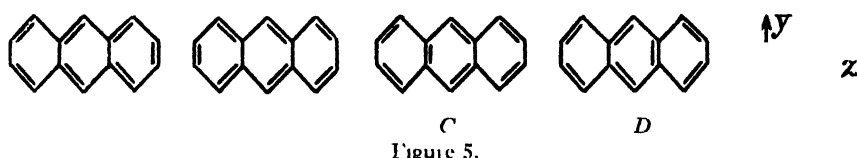


Figure 5.

is clearly polarized in the z -direction, being $A_{1g} \rightarrow B_{1u}$. This is contrary to our own conclusions. But two comments must be made. First, all higher excited structures were omitted. Now both Jonsson (1942) and A. and B. Pullman (1946) have shown that these are very significant, even in the ground state, contributing at least 50%, additional resonance energy, and having a total weight almost equal to that of the four Kekulé structures. Second, the very nature of the four wave functions $A \dots D$ precludes the possibility of any charge transfer in the y -direction. To get this we must include ionic structures, of which two types (cf. Jones 1945) are likely to be most important in determining intensities. We show (E - F , G - H) two examples of each, though there are many more from which we might have made our selection. E and F may be called y -ionic structures since they will

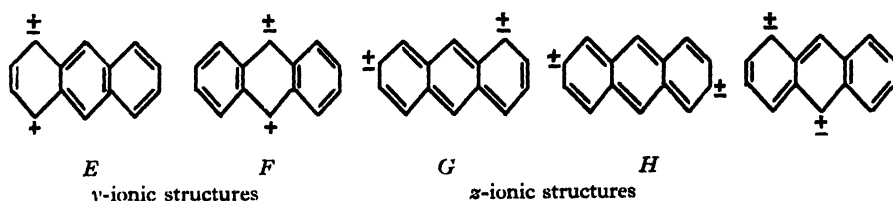


Figure 6.

contribute to transitions polarized in that direction: G and H contribute mostly to z -polarization. There are, however, structures such as I which contribute to both polarizations, and which are not unfavourable energetically. Now a careful consideration of the ionic structures E - H shows that in the y -ionic type the resonance in at least one ring, and often in two rings, is unbroken, whereas in the z -ionic type, resonance is broken in every ring, the structures being essentially

Excited Electronic Levels in Conjugated Molecules: II. On the Symmetry and Multiplicity of Molecular States

By H. C. LONGUET-HIGGINS

Physical Chemistry Laboratory, Oxford

Communicated by C. A. Coulson; MS. received 28 June 1947

ABSTRACT. A comparison of the valence-bond method and the method of molecular orbitals shows that the valence-bond method, while often more reliable for the calculation of ground-state energies, leads to erroneous conclusions as to the number and nature of the excited states, whereas the method of molecular orbitals gives the right number of excited states of any specified symmetry and multiplicity, and leads to semi-quantitatively correct energies for the ground state and lowest excited states. The latter method therefore seems to be the more reliable for the interpretation of electronic absorption spectra in conjugated systems.

§ 1. INTRODUCTION

IT is well known that there are many molecules for which the two standard approximate methods of solving the wave equation, namely the valence-bond method and the method of molecular orbitals, lead to different results for the multiplicities and symmetries of the ground state and of the first, second, and subsequent excited states, and even lead to different results for the number of molecular states of given multiplicity and symmetry.

In this paper we shall enquire into the origin of these discrepancies, and shall attempt to assess the relative merits and defects of the two approaches in the interpretation of molecular spectra.

§ 2. THE HYDROGEN MOLECULE*

We shall first deal briefly with the hydrogen molecule, as it illustrates simply certain general points that we shall make later on. We assume the wave function of any molecular state to be expressible as a sum of products of $1s$ atomic wave functions for the atoms composing the molecule (AO approximation). Let us denote the $1s$ atomic wave function for an electron of spin $\frac{1}{2}$ on nucleus A by a , and the $1s$ wave function for an electron of spin $-\frac{1}{2}$ on the same nucleus by \bar{a} . b , \bar{b} will denote atomic wave functions for electrons of spin $\frac{1}{2}$, $-\frac{1}{2}$ on nucleus B. Further, we shall adopt the convention that the first factor in a product of atomic wave functions is a function of the coordinates of electron 1, and the second factor a function of the coordinates of electron 2. Thus, $b\bar{a}$ denotes the wave function for a hypothetical state in which electron 1 is on nucleus B with spin $\frac{1}{2}$, and electron 2 on nucleus A with spin $-\frac{1}{2}$.

Now, by Pauli's exclusion principle, the total molecular wave function must change sign when we interchange the two electrons. Therefore in the AO approximation the wave function for any molecular state must be some linear combination of the terms

$$ab - ba, \quad a\bar{b} - \bar{b}a, \quad b\bar{a} - \bar{a}b, \quad \bar{a}\bar{b} - \bar{b}\bar{a}, \quad a\bar{a} - \bar{a}a \quad \text{and} \quad b\bar{b} - \bar{b}b.$$

* The argument of this section follows closely the lines of a discussion given by Hund (1932), and quoted by Van Vleck and Sherman (1935).

Of these the first four are called *covalent terms*, and the other two *ionic terms*. Further, any wave function for the molecule must have the symmetry of some irreducible representation of the molecular group; for H_2 this means that the wave function must be either symmetric or antisymmetric with respect to exchange of the nuclei.

It follows from this that the possible molecular wave functions are, apart from normalization factors:—

$$\text{I} \quad \lambda_1 \{ (a\bar{b} - \bar{b}a) + (b\bar{a} - \bar{a}b) \} + \mu_1 \{ (a\bar{a} - \bar{a}a) + (b\bar{b} - \bar{b}b) \} \quad S_N$$

$$\text{II} \quad \lambda_2 \{ (a\bar{b} - \bar{b}a) + (b\bar{a} - \bar{a}b) \} + \mu_2 \{ (a\bar{a} - \bar{a}a) + (b\bar{b} - \bar{b}b) \} \quad S_N$$

where $\lambda_1 : \mu_1$ and $\lambda_2 : \mu_2$ may be determined by a variation method

$$\text{III} \quad (a\bar{a} - \bar{a}a) - (b\bar{b} - \bar{b}b) \quad A_N$$

$$\text{IV (i)} \quad (ab - ba) \quad A_N$$

$$\text{(ii)} \quad (a\bar{b} - \bar{b}a) - (b\bar{a} - \bar{a}b) \quad A_N$$

$$\text{(iii)} \quad (\bar{a}\bar{b} - \bar{b}\bar{a}) \quad A_N$$

The symbols S_N , A_N denote whether the wave function is symmetric or antisymmetric with respect to interchange of the nuclei.

The above wave functions may all be factorized into an orbital part and a spin part. When this is done it turns out that I, II and III all contain the spin function $\alpha(1)\beta(2) - \beta(1)\alpha(2)$, and have different orbital functions, whereas the wave functions IV contain the same orbital function multiplying three different spin functions, namely $\alpha(1)\alpha(2)$, $\beta(1)\beta(2)$ and $\alpha(1)\beta(2) + \beta(1)\alpha(2)$. I, II and III are therefore singlet levels, and IV is a triplet level.*

In the simple valence-bond picture the terms $a\bar{a}$, $b\bar{b}$ are ignored, so that we are left with the molecular wave functions:

$$\text{V} \quad (a\bar{b} - \bar{b}a) + (b\bar{a} - \bar{a}b) \quad S_N$$

$$\text{VI (i)} \quad (ab - ba) \quad A_N$$

$$\text{(ii)} \quad (a\bar{b} - \bar{b}a) - (b\bar{a} - \bar{a}b) \quad A_N$$

$$\text{(iii)} \quad (\bar{a}\bar{b} - \bar{b}\bar{a}) \quad A_N$$

Of these the first is obtained by putting $\mu_1 = 0$ in I, or $\mu_2 = 0$ in II; and the others are the same as IV. We have thus discarded the A_N singlet state and one of the S_N singlet states.

In the molecular-orbital treatment we construct the molecular wave functions out of wave functions u and v , each of which represents a possible wave function for a single electron in the field of both nuclei. It is clear that u and v must each be symmetric or antisymmetric with respect to exchange of the nuclei, so that $u = a + b$, $v = a - b$.

* The conventional notation for classifying molecular wave functions compounded of atomic orbitals of the s type is as follows:—

Multiplicity	Behaviour with respect to interchange of the nuclei	Symbol
Singlet	Symmetric	$^1\Sigma_g$
Singlet	Antisymmetric	$^1\Sigma_u$
Triplet	Symmetric	$^3\Sigma_g$
Triplet	Antisymmetric	$^3\Sigma_u$

The possible molecular wave functions are then :

VII	$(u\bar{u} - \bar{u}u)$	S_N
VIII	$(v\bar{v} - \bar{v}v)$	S_N
IX	$(u\bar{v} - \bar{v}u) + (v\bar{u} - \bar{u}v)$	A_N
X (i)	$(uv - vu)$	A_N
(ii)	$(u\bar{v} - \bar{v}u) - (v\bar{u} - \bar{u}v)$	A_N
(iii)	$(\bar{u}\bar{v} - \bar{v}\bar{u})$.	

Substitution of $a + b$ for u , and $a - b$ for v , gives :

VII	$(a\bar{a} - \bar{a}a) + (b\bar{b} - \bar{b}b) + (a\bar{b} - \bar{b}a) + (b\bar{a} - \bar{a}b)$	S_N
VIII	$(a\bar{a} - \bar{a}a) + (b\bar{b} - \bar{b}b) - (a\bar{b} - \bar{b}a) - (b\bar{a} - \bar{a}b)$	S_N
IX	$2(a\bar{a} - \bar{a}a) - 2(b\bar{b} - \bar{b}b)$	A_N
X (i)	$-2(ab - ba)$	A_N
(ii)	$-2(a\bar{b} - \bar{b}a) + 2(b\bar{a} - \bar{a}b)$	A_N
(iii)	$-2(\bar{a}\bar{b} - \bar{b}\bar{a})$	A_N

We see at once that, when account has been taken of normalization, the functions X are identical with IV, and that IX is identical with III. VII and VIII are simply linear combinations of I and II, and vice versa. The molecular-orbital method therefore gives us the right number of molecular states, with the right symmetry and multiplicity. On the other hand, if we take I as representing the ground state, calculation shows that $\lambda_1 \gg \mu_1$, so that V gives a better representation of the ground state than VII.

It is well known that if we take into account the ionic terms in the valence-bond method, the wave functions I and II that we shall get by a variational treatment will be as good as can be obtained on the *AO* approximation. What is not so commonly realized is that by taking into account interactions between the wave functions VII and VIII we can obtain the same best-possible wave functions I and II; that is to say, at the next degree of approximation the two methods converge.

We have considered the familiar case of H_2 in some detail because it exemplifies the results that will be established later : namely that in the valence-bond method a quite simplified treatment will give perhaps a better value for the energy in the ground state, but leads to errors of omission in the excited states ; whereas the method of antisymmetrized molecular orbitals, though it may not give such an accurate wave function for the ground state, always leads to qualitatively correct conclusions about the number, symmetry and multiplicity of the excited states. In principle, of course, this defect in the valence-bond method could be eliminated by the inclusion of all possible ionic terms, but this is not feasible in complex molecules.

§ 3. MORE COMPLEX MOLECULES

We shall now consider the relation between the valence-bond and molecular-orbital methods in more complex molecules: In particular we shall study conjugated systems, that is, sets of more than two overlapping atomic orbitals.

These are well exemplified, of course, by the poly-unsaturated hydrocarbons and their derivatives, where the π -electrons move in overlapping sets of dumb-bell-shaped atomic orbitals.

Let us denote by a, b, c, \dots the atomic wave functions for a single electron of spin $\frac{1}{2}$ on atoms A, B, C, ... of a conjugated system, and by $\bar{a}, \bar{b}, \bar{c}, \dots$ the corresponding wave functions when the electron has spin $-\frac{1}{2}$. Let us again adopt the convention that the first term in a product of these atomic orbitals is a function of the coordinates of electron 1, the second term a function of the coordinates of electron 2, and so on. Then in the AO approximation, using the exclusion principle, we write the molecular wave function as a sum of terms of the type

$$\Sigma P . a\bar{b}\bar{c}d\bar{d},$$

where ΣP denotes the sum of the even permutations of $a\bar{b}\bar{c}d\bar{d}$ minus the sum of the odd permutations. We shall call such an expression an *atomic term*. For example, the atomic term $\Sigma P . a\bar{b}\bar{c}$, when written out in full, becomes

$$a(1)b(2)\bar{c}(3) + b(1)\bar{c}(2)a(3) + \bar{c}(1)a(2)b(3) - a(1)\bar{c}(2)b(3) - \bar{c}(1)b(2)a(3) - b(1)a(2)\bar{c}(3).$$

It is clear that such a sum will vanish if the same atomic orbital (a.o.) occurs twice in the sum and with the same spin, which means physically that two electrons cannot occupy the same orbital with the same spin. We may therefore classify our atomic terms according to the number of atomic orbitals that are doubly occupied in the corresponding configuration. For instance, in a system of 4 electrons and 4 atomic orbitals we have:

- (i) covalent terms, e. g. $\Sigma P . a\bar{b}\bar{c}d$,
- (ii) singly ionic terms, e. g. $\Sigma P . a\bar{a}b\bar{c}$,
- (iii) doubly ionic terms, e. g. $\Sigma P . a\bar{a}\bar{c}\bar{c}$.

In general, if there are $2n$ electrons and $2n - p$ atomic orbitals, we shall have

$$\frac{(2n-p)! 2^{2n-2r}}{r!(r-p)!(2n-2r)!} \quad r\text{-ply ionic terms, } (r \geq p),$$

because there are

$$\frac{(2n-p)!}{r!(r-p)!(2n-2r)!}$$

ways of choosing r doubly occupied a.o. and $r-p$ unoccupied ones, and we may then assign spins to the remaining singly occupied a.o. in 2^{2n-2r} different ways. The total number of terms to be considered is therefore

$$\sum_r \frac{(2n-p)! 2^{2n-2r}}{r!(r-p)!(2n-2r)!}.$$

We could determine the wave functions and energies of the various molecular states by setting up the secular equation in terms of all the atomic terms, and solving it by straightforward methods. However, before doing this it is best to group the atomic terms corresponding to any given distribution of electrons among the a.o. into certain combinations such that the secular equation automatically factorizes into smaller equations when set up in terms of these combinations, each of the smaller equations giving the energies and wave functions of the molecular states of one particular multiplicity. We shall call these special combinations of

atomic terms *spin-grouped atomic terms*. Any one spin-grouped atomic term will only contribute to the molecular states of one particular multiplicity, and may therefore be designated as a singlet, triplet etc., atomic term. For instance, in a molecule with 4 electrons and 4 a.o., the following will be typical spin-grouped atomic terms:

$$\Sigma P . (\bar{a}\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}\bar{c}\bar{d}), \quad \Sigma P . (\bar{a}\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}\bar{c}\bar{d}), \quad \Sigma P . (\bar{a}\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}\bar{c}\bar{d}),$$

which contribute to the triplet states, and

$$\Sigma P . (\bar{a}\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}\bar{c}\bar{d} + \bar{a}\bar{b}\bar{c}\bar{d}), \quad \Sigma P . (\bar{a}\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}\bar{c}\bar{d} + \bar{a}\bar{b}\bar{c}\bar{d}),$$

which contribute to the singlet levels.

Now the molecular wave functions belonging to a degenerate molecular state will differ in total electronic spin; and it can be shown that every degenerate molecular state has just one wave function of zero spin belonging to it. For instance, for the triplet level of H_2 the wave function of zero spin is

$$(\bar{a}\bar{b} - \bar{b}\bar{a}) - (\bar{h}\bar{a} - \bar{a}\bar{h}).$$

Therefore the total number of molecular states of multiplicity $2q-1$ is equal to the total number of spin-grouped atomic terms of the same multiplicity and zero total spin. Now, corresponding to a given r -ply ionic distribution of electrons among the atoms of our system, there are $\binom{2n-2r}{n-r}$ different r -ply ionic terms with zero total spin. From these we can construct

$$\binom{2n-2r}{n-r} - \binom{2n-2r}{n-r-1} \text{ singlet spin-grouped atomic terms,}$$

$$\binom{2n-2r}{n-r-1} - \binom{2n-2r}{n-r-2} \text{ triplet terms,}$$

etc., and one $(2n-2r+1)$ -plet term. (Eckart 1930)

Therefore the total number of spin-grouped terms of multiplicity $2q-1$ which can be constructed out of r -ply ionic terms is

$$r!(r-p)!(2n-2r)! \left\{ \binom{2n-2r}{n-r-q+1} - \binom{2n-2r}{n-r-q} \right\} \\ - \frac{(2n-p)!(2q-1)}{r!(r-p)!(n-r-q+1)!(n-r+q)!} \text{ when } p < r < n-q,$$

or

$$\frac{(2n-p)!}{r!(r-p)!(2n-2r)!} = \frac{(2n-p)!}{(n-q+1)!(n-q-p+1)!(2q-2)!},$$

when $r=n-q+1$, or 0 when $r > n-q+1$.

The total number of molecular wave functions of multiplicity $2q-1$ is therefore

$$\sum_{r=p}^{n-q} \frac{(2n-p)!(2q-1)}{r!(r-p)!(n-r-q+1)!(n-r+q)!} + \frac{(2n-p)!}{(n-q+1)!(n-q-p+1)!(2q-2)!}$$

when $q \leq n-p$, and 0 otherwise.

This is the number we should obtain by the valence-bond method taking into account up to s -fold ionic terms, provided $s \geq n-q+1$. If s were less than

$n - q + 1$, we should get only some of the molecular states of multiplicity $2q - 1$; and in particular we shall only get the correct number of singlet states by taking account of all possible ionic terms ($s = n$).

In the molecular orbital approximation we express the molecular wave function in terms of single-electron wave functions, each of which is a linear combination of the atomic orbitals in the system. These single-electron wave functions are called molecular orbitals (m.o.), and are equal in number to the component atomic orbitals.

Let us denote the molecular orbitals for an electron of spin $\frac{1}{2}$ by u, v, w, \dots and those for an electron of spin $-\frac{1}{2}$ by $\bar{u}, \bar{v}, \bar{w}, \dots$. We shall continue in the convention that the first factor in a product of single-electron functions is a function of the coordinates of the first electron, the second a function of the coordinates of the second electron, and so on.

Then using the exclusion principle we write the molecular wave function as a sum of terms of the type $\Sigma P \cdot uv\bar{w}x$, where the operator ΣP has the same significance as before. We call such an expression a *molecular term*. A molecular term will vanish if the same m.o. occurs in it twice with the same spin, so two electrons with the same spin cannot occupy the same m.o. Just as for atomic terms we may classify our molecular terms according to their numbers of doubly filled m.o. Thus if we have a system of 4 electrons and 4 atomic orbitals, and hence 4 molecular orbitals, then there will be three kinds of molecular term, viz:

- (i) uncoupled terms, e.g. $\Sigma P \cdot uv\bar{w}x$,
- (ii) singly coupled terms, e.g. $\Sigma P \cdot u\bar{u}v\bar{v}$,
- (iii) doubly coupled terms, e.g. $\Sigma P \cdot u\bar{u}v\bar{v}$.

There is a complete analogy between r -ply coupled molecular terms and r -ply ionic atomic terms, as we shall see at once. If our system has $2n$ electrons and $2n - p$ m.o. (compounded of the same number of a.o.), then there will be

$$\frac{(2n - p)! 2^{2n - 2r}}{r! (r - p)! (2n - 2r)!} r\text{-ply coupled molecular terms } (r \geq p),$$

since there are

$$\frac{(2n - p)!}{r! (r - p)! (2n - 2r)!}$$

ways of choosing r m.o. to be doubly filled and $r - p$ to be unfilled, and we may then assign spins to the remaining m.o. in $2^{2n - 2r}$ different ways. The total number of molecular terms to be considered is therefore

$$\sum_{r=p}^n \frac{(2n - p)! 2^{2n - 2r}}{r! (r - p)! (2n - 2r)!}.$$

From these molecular terms one can construct certain linear combinations—"spin-grouped molecular terms"—analogous to spin-grouped atomic terms, such that the secular equation would factorize into smaller equations if set up in terms of them, each smaller equation giving the energies and wave functions of the molecular states of one particular multiplicity. Any one spin-grouped term will only contribute to the molecular states of one particular multiplicity, and may therefore be designated as a singlet, triplet, $(2q - 1)$ -plet molecular term. The construction of

spin-grouped molecular terms is precisely analogous to the construction of spin-grouped atomic terms. For example, to get a set of linearly independent singlet molecular terms we may write down the m.o. in a circle and apply Rumer's theorem. Thus if we have 4 electrons and 4 m.o. u, v, w and x , the non-crossing diagrams

and



will represent the singlet terms

$$\Sigma P . (u\bar{x} - \bar{u}x)(v\bar{w} - \bar{v}w) \quad \text{and} \quad \Sigma P . (u\bar{v} - \bar{u}v)(w\bar{x} - \bar{w}x).$$

The total number of spin-grouped molecular terms of multiplicity $2q-1$ and zero spin which can be constructed out of r -ply coupled molecular terms is

$$\frac{(2n-p)!}{r!(r-p)!(2n-2r)!} \left\{ \binom{2n-2r}{n-r-q+1} - \binom{2n-2r}{n-r-q} \right\} \\ = \frac{(2n-p)!(2q-1)}{r!(r-p)!(n-r-q+1)!(n-r+q)!} \quad \text{when } p \leq r \leq n-q,$$

or

$$\frac{(2n-p)!}{r!(r-p)!(2n-2r)!} = \frac{(2n-p)!}{(n-q+1)!(n-q-p+1)!(2q-2)!}$$

when $r = n - q + 1$, and zero otherwise. But the total number of $(2q-1)$ -plet molecular states must be equal to the total number of $(2q-1)$ -plet molecular terms of zero spin, and this number is

$$\sum_p^{n-q} \frac{(2n-p)!(2q-1)}{r!(r-p)!(n-r-q+1)!(n-r+q)!} + \frac{(2n-p)!}{(n-q+1)!(n-q-p+1)!(2q-2)!}$$

when $q \leq n-p$, and 0 otherwise. Therefore whether we express our molecular wave functions as combinations of molecular or atomic terms, we get the same number of $(2q-1)$ -plet molecular states. This must necessarily be so, since the $(2q-1)$ -plet molecular terms are linear combinations of $(2q-1)$ -plet atomic terms, and so whichever we use as basis for solving the secular equation we shall get the same final molecular wave functions.

Having obtained our spin-grouped molecular terms we might then solve the secular equations in terms of them. However, this would give us the same results as a valence-bond treatment taking account of all ionic terms, and with even greater labour. So in the method of m.o. as ordinarily applied we observe that the wave functions for the $(2q-1)$ -plet molecular states are equal in number to the $(2q-1)$ -plet molecular terms, and we identify the latter with the former. This short cut is valid so long as the energies of the various spin-grouped molecular terms differ considerably, since then the interactions between them will be small; but if the molecule has symmetry, some additional considerations are necessary.

§ 4. SYMMETRY OF MOLECULAR STATES

If the molecule has spatial symmetry, any actual wave function for the molecule must have the symmetry of some irreducible representation of the molecular group. It follows that the linear combinations of atomic or molecular terms by which we represent the wave function must also have such symmetry. Now an atomic term does not in general have irreducible symmetry, being a product of arbitrarily

selected atomic wave functions ; but a molecular term, being a product of molecular orbitals each of which has irreducible symmetry, usually does have irreducible symmetry. In molecules whose symmetry groups contain degenerate representations, the molecular terms will not all have irreducible symmetry, but very simple combinations of them can be found which do ; we shall ignore this case for the moment. Now a molecular wave function of given multiplicity and irreducible symmetry must be compounded of spin-grouped molecular terms of the same symmetry and multiplicity. Hence the number of spin-grouped molecular terms of given symmetry and multiplicity and zero total spin is equal to the number of molecular states of the same symmetry and multiplicity.

In the method of antisymmetrized m.o., as mentioned in the last paragraph, it is customary to identify the spin-grouped molecular terms with the actual molecular wave functions (in molecules whose symmetry groups contain no degenerate representations). We found that this leads to the correct number of singlet, triplet, etc., molecular states, and we now see that it gives us the correct number of singlet, triplet, etc., molecular states of given irreducible symmetry. We also saw in the last paragraph that this short cut is equivalent to neglecting the interactions between different spin-grouped molecular terms. When the molecule is symmetrical there can be no interactions between terms of different symmetry, and so we only have to consider whether terms of the same symmetry and multiplicity are going to interact appreciably. If their energies are close together there will be considerable interaction, but it often happens that this is not so ; for instance in H_2 the energies of the molecular terms VII and VIII differ considerably, since the former corresponds to a situation in which both electrons are in the bonding m.o. $a + b$, and the latter to both electrons being in the antibonding m.o. $a - b$.

In the valence-bond method, the spin-grouped ionic terms are frequently referred to as "excited states". This nomenclature is misleading, as it implies that they approximate to the wave functions for actual molecular states, which is not generally the case, for two reasons. First, if the molecule is symmetrical, any arbitrarily selected atomic term will in general be asymmetric ; it is only selected combinations of similar terms that will have irreducible symmetry in the molecular group. Secondly, the energies of these selected combinations will be roughly equal for terms with the same ionicity, so that the interactions between combinations with the same symmetry and multiplicity are by no means negligible. It is perhaps too much to hope that the use of the term "excited states" in this connection will disappear altogether ; but it would be a step forward if its use could be confined at least to irreducibly symmetric *and* spin-grouped combinations of atomic terms.

There is one point which has been passed over in the above treatment. If we are treating a molecule by the method of antisymmetrized m.o., then, as mentioned above, we identify the actual molecular states with spin-grouped molecular terms ; this is equivalent to saying that in a given molecular state such and such m.o. are singly or doubly occupied, and the multiplicity is $2q - 1$, say. This procedure gives an unambiguous wave function for the given state provided that not more than two m.o. are singly occupied ; for instance, the wave function for the singlet state u^2v^2 is $\Sigma P . u\bar{u}v\bar{v}$, and those for the triplet state u^2vw are $\Sigma P . u\bar{u}(v\bar{w} + \bar{v}w)$, $\Sigma P . u\bar{u}v\bar{w}$ and $\Sigma P . u\bar{u}\bar{v}w$. On the other hand, all we know about the wave functions for the singlet states $uvwx$ is that they are two orthogonal linear combinations of $\Sigma P . (u\bar{v} - \bar{u}v)(w\bar{x} - \bar{w}x)$ and $\Sigma P . (u\bar{x} - \bar{u}x)(v\bar{w} - \bar{v}w)$. To find the

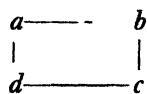
correct combinations we should have to solve the secular equation, which necessitates knowing the interactions between the m.o. Fortunately, however, the important excited states of a molecule are generally the lowest ones, in which not more than two electrons are simultaneously unpaired, so that there is no ambiguity about the form of the wave function.

§ 5. AN EXAMPLE: 4 ELECTRONS IN 4 ATOMIC ORBITALS

It will probably clarify the above argument if we consider in detail a hypothetical molecule consisting of 4 atoms arranged at the corners of a rectangle. Such a molecule is the prototype of naphthalene and anthracene, which belong to the same symmetry group D_{2h} , and so our general conclusions will apply to these molecules also. Let us assume that the atomic orbitals in our rectangular molecule are antisymmetric with respect to the plane of the molecule, like the $2p_z$ orbitals in which π -electrons move. Then it will be impossible to construct molecular wave functions that are anything but antisymmetric with respect to the molecular plane, and so we shall lose nothing by taking the symmetry group of the molecule to be the simpler group C_{2v} . The character table of the group C_{2v} is as follows:—

	E	C_2	σ_x	σ_y
A_1	1	1	+1	+1
A_2	1	1	-1	-1
B_1	1	-1	+1	-1
B_2	+1	-1	-1	+1

We will represent the molecule diagrammatically thus:



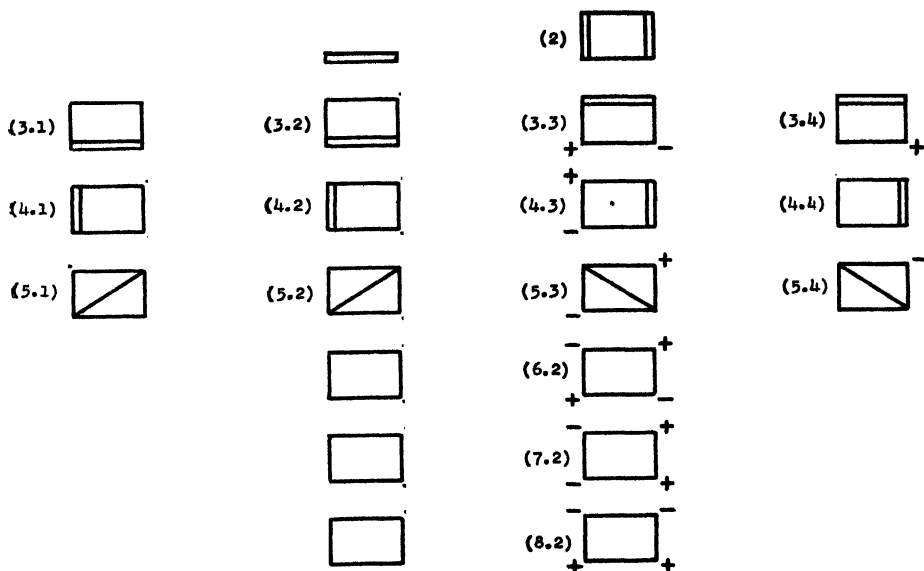
where the bonds lengths are not all equal, but $x_{ab} = x_{cd}$ and $x_{bc} = x_{da}$.

Let us first take the valence-bond approach. Then there are three kinds of atomic term—covalent, singly ionic and doubly ionic. These may be combined into singlet, triplet and quintuplet spin-grouped atomic terms. The singlet terms may be represented by structures of the conventional type; for instance (1) represents $\Sigma P . (\bar{a}\bar{b} - \bar{a}b)(\bar{c}\bar{d} - \bar{c}d)$, (3.1) represents $\Sigma P . \bar{b}\bar{b}(\bar{c}\bar{d} - \bar{c}d)$ and (8.1) represents $\Sigma P . \bar{c}\bar{c}\bar{d}\bar{d}$. The following diagram gives all such structures which are possible. Of these singlet atomic terms only the first two have irreducible symmetry. As for the others, we can group congruent structures into combinations having irreducible symmetry: for example, the combination (7.1) + (7.2) belongs to the irreducible representation A_1 , and (7.1) - (7.2) belongs to A_2 . There will be as many of these irreducibly symmetric combinations of a set of congruent atomic terms as there are atomic terms in the set.

We may draw up a table giving the number of irreducibly symmetric combinations of atomic terms in each irreducible representation. The following is such a table:—

	A_1	A_2	B_1	B_2
(1)	1			
(2)	1			
(3.1) to (3.4)	1	1	1	1
(4.1) to (4.4)	1	1	1	1
(5.1) to (5.4)	1	1	1	1
(6.1) and (6.2)	1	1		
(7.1) and (7.2)	1		1	
(8.1) and (8.2)	1			1
Total	8	4	4	4 = 20

Since the molecular wave functions are built up of these symmetric combinations, there must be eight singlet molecular wave functions of symmetry A_1 , and four in each of the representations A_2 , B_1 and B_2 . If we had ignored the doubly ionic terms (6.1) to (8.2), we should only have found five singlet states of symmetry A_1 , three of symmetry A_2 , three of symmetry B_1 , and three of symmetry B_2 . That is, we should have missed six possible singlet molecular states.



List of singlet structures in rectangular molecule.

The triplet states may be discussed in a similar way. There are altogether fifteen triplet atomic terms of zero spin, namely the following:—

$$\begin{array}{llll}
 \Sigma P . (a\bar{b}\bar{c}\bar{d} - \bar{a}\bar{b}cd), & \Sigma P . (a\bar{b}\bar{c}d - \bar{a}\bar{b}cd), & \Sigma P . (a\bar{b}\bar{c}d - \bar{a}\bar{b}cd), & \\
 (9.1) & (9.2) & (9.3) & \\
 \Sigma P . b\bar{b}(c\bar{d} + \bar{c}d), & \Sigma P . a\bar{a}(c\bar{d} + \bar{c}d), & \Sigma P . c\bar{c}(a\bar{b} + \bar{a}b), & \Sigma P . d\bar{d}(a\bar{b} + \bar{a}b), \\
 (10.1) & (10.2) & (10.3) & (10.4) \\
 \Sigma P . c\bar{c}(d\bar{a} + \bar{d}a), & \Sigma P . b\bar{b}(d\bar{a} + \bar{d}a), & \Sigma P . d\bar{d}(b\bar{c} + \bar{b}c), & \Sigma P . a\bar{a}(b\bar{c} + \bar{b}c), \\
 (11.1) & (11.2) & (11.3) & (11.4) \\
 \Sigma P . a\bar{a}(b\bar{d} + \bar{b}d), & \Sigma P . c\bar{c}(d\bar{b} + \bar{d}b), & \Sigma P . b\bar{b}(c\bar{a} + \bar{c}a), & \Sigma P . d\bar{d}(a\bar{c} + \bar{a}c). \\
 (12.1) & (12.2) & (12.3) & (12.4)
 \end{array}$$

It is probably misleading to represent these by bond diagrams, as the symmetry is obscured. As before, we may group these triplet terms into irreducibly symmetrical combinations. For instance, (9.2) belongs to A_2 , (9.1)+(9.3) to B_1 , and (9.1)–(9.3) to B_2 . The following gives the number of combinations in each representation:—

	A_1	A_2	B_1	B_2				
(9.2)		1						
(9.1) and (9.3)			1	1				
(10.1) to (10.4)	1	1	1	1				
(11.1) to (11.4)	1	1	1	1				
(12.1) to (12.4)	1	1	1	1				
Total	3	+	4	+	4	+	4	=15

There are therefore three triplet molecular states of irreducible symmetry A_1 and four in each of the symmetry classes A_2 , B_1 and B_2 . Since the doubly ionic terms contribute nothing to the triplet levels, we should not miss any triplet states by neglecting them; but we should miss all except three of the triplet states if we neglected singly ionic terms as well.

There is only one quintuplet spin-corrected atomic term of zero spin, namely

$$\Sigma P . (abc\bar{d} + a\bar{b}c\bar{d} + a\bar{b}c\bar{d} + \bar{a}bcd + \bar{a}bcd + \bar{a}bcd).$$

This has symmetry A_1 , and as there is no other quintuplet with which it can interact, it must be the molecular wave function for the quintuplet level of zero spin.

Now let us attack the problem by the method of molecular orbitals. There will be one m.o. in each symmetry class, namely

$$(a + b + c + d) = a_1, \text{ say, in class } A_1,$$

$$(a - b + c - d) = a_2, \text{ say, in class } A_2,$$

$$(a - b - c + d) = b_1, \text{ say, in class } B_1,$$

$$(a + b - c - d) = b_2, \text{ say, in class } B_2.$$

The possible molecular terms are of three types: doubly coupled, e.g. $\Sigma P . a_1\bar{a}_1b_1\bar{b}_1$, singly coupled, e.g. $\Sigma P . a_1\bar{a}_1a_2\bar{b}_1$, and uncoupled, e.g. $\Sigma P . a_1\bar{a}_2\bar{b}_1b_2$. Each of these has irreducible symmetry, as the group C_{2v} has no degenerate representations. Proceeding as usual, we group the molecular terms of zero total spin into spin-grouped terms. For example,

$\Sigma P . (a_1\bar{a}_2 - \bar{a}_1a_2)(b_1\bar{b}_2 - b_1\bar{b}_2)$ is a singlet molecular term,

$\Sigma P . a_1\bar{a}_1(b_1\bar{b}_2 + \bar{b}_1b_2)$ is a triplet molecular term, and

$\Sigma P . (a_1a_2\bar{b}_1\bar{b}_2 + a_1\bar{a}_2b_1\bar{b}_2 + a_1\bar{a}_2\bar{b}_1b_2 + \bar{a}_1a_2b_1\bar{b}_2 + \bar{a}_1\bar{a}_2\bar{b}_1b_2 + \bar{a}_1\bar{a}_2b_1b_2)$ is the only quintuplet term.

We may draw up a table of possible distributions of electrons among the m.o., giving the class of the distribution and the number of levels of given multiplicity belonging to it. A doubly coupled distribution gives rise to one singlet term, a singly coupled distribution to one singlet and one triplet, and the uncoupled distribution $a_1a_2b_1b_2$ gives rise to 2 singlets, three triplets and one quintuplet term. We identify these spin-corrected terms with the actual molecular wave functions.

Distribution of electrons among m.o.				Symmetry class	Number of levels		
a_1	a_2	b_1	b_2		Singlet	Triplet	Quintuplet
2	2			A_1	1		
2	1	1		B_2	1	1	
2	1		1	B_1	1	1	
2		1	1	A_2	1	1	
2		2		A_1	1		
2			2	A_1	1		
1	2	1		B_1	1	1	
1	2		1	B_2	1	1	
1	1	2		A_2	1	1	
1	1	1	1	A_1	2	3	1
1		2	1	B_2	1	1	
1		1	2	B_1	1	1	
	2	2		A_1	1		
	2	1	1	A_2	1	1	
	2		2	A_1	1		
	1	2	1	B_1	1	1	
	1	1	2	B_2	1	1	
		2	2	A_1	1		
1	1		2	A_2	1	1	

It will be observed that the number of levels of given symmetry and multiplicity is exactly the same as is obtained by the complete valence-bond treatment. In particular it can be verified by expansion that, apart from a normalization factor, the quintuplet level obtained by the m.o. method is identical with that obtained by the valence-bond method, and contains no ionic terms although it can be expressed entirely in terms of molecular orbitals. Hence the method of m.o. does not always exaggerate the importance of ionic terms.

One further point should be noted. If the energies of the m.o., a_1 , a_2 , b_1 , b_2 are nearly in arithmetic progression, the excited states (singlet or triplet) $a_1^2 a_2 b_2$ and $a_1 a_2^2 b_1$, being of the same symmetry and multiplicity, and about the same energy, will interact; and the states represented by the descriptions $a_1^2 b_2^2$, $a_1 a_2 b_1 b_2$ and $a_2^2 b_1^2$ will also interact, so that the molecular orbital method will give inaccurate energies for the resulting molecular states. However, in the ground state and first excited state this difficulty does not arise; and in the absorption of light by unsaturated molecules it is the ground state and usually the first excited state between which the visible transition occurs.

§6. CONCLUSION

We may summarize the situation as follows:—

(1) Although neither the valence-bond method nor molecular orbital method as usually applied gives an unequivocal answer to the question of the symmetry and multiplicity of states of progressively increasing energy (see Wheland 1938), the latter method will always give the correct *number* of molecular states of given symmetry and multiplicity, whereas the former method will only do so if all possible ionic terms are taken into account.

(2) As to the quantitative merits of the two theories, it seems fair to say that the valence-bond method neglecting some or all ionic states will probably give a more accurate value for the energy of the ground state when this is a singlet, but that it will be unreliable for estimating the energies of specified excited states, partly owing to the difficulty of deciding which actual molecular state is being discussed. On the other hand, the method of m.o. will lead to fairly reliable values of the energies of ground state and first one or two excited states (so long as there are no other states of the same symmetry and multiplicity and nearly the same energy as the state under consideration). If interaction between m.o. molecular wave functions of the same symmetry and multiplicity is considered, the resulting molecular wave functions and energies will be identical with those obtained by a valence-bond treatment with all ionic terms taken into account. In practice either of these last procedures is excessively laborious.

(3) The method of molecular orbitals therefore seems to be the more reliable for the interpretation of what Mulliken calls $N \rightarrow V$ type molecular spectra, where the most important lines are due to transitions between the ground state and lower excited states. For Rydberg type transitions in electronic spectra, the AO approximation breaks down, and so either method is inappropriate.

REFERENCES

- ECKART, C., 1930, *Rev. Mod. Phys.*, **2**, 366–370.
 HUND, F., 1932, *Z. Phys.*, **73**, 1.
 VAN VLECK, J. H., and SHERMAN, A., *Rev. Mod. Phys.*, **7**, 183.
 WHELAND, G. W., 1938, *Proc. Roy. Soc. A*, **164**, 397.

The Measurement of the Permeability of Low-conductivity Ferromagnetic Materials at Centimetre Wavelengths

By J. B. BIRKS

I.C.I. Research Fellow, University of Glasgow

MS. received 21 April 1947

ABSTRACT. A waveguide impedance method for the measurement of the permeability of low-conductivity ferromagnetic materials at centimetre wavelengths is described. The method avoids the difficulties and anomalies present in earlier measurements. The complex magnetic and dielectric properties of a specimen filling a section of waveguide are derived from standing-wave measurements of the short-circuit and open-circuit impedances of the section. The experimental apparatus and techniques which have been employed at wavelengths of 60–20 cm., 15.6 cm., 3 cm. and 1.4 cm. are described, and some typical results of measurements on γ -ferric oxide are given. It is shown that the observed magnetic dispersion is an inherent property of the material, and is not attributable to skin-effect.

§ 1. INTRODUCTION

ALMOST all the previous experimental work on high-frequency permeability, reviewed by Allanson (1945), has been concerned with ferromagnetic metals. These materials, because of their high conductivity, may be studied as conducting elements in a high-frequency transmission line, or in a cavity resonator (Griffiths 1946). The experimental values of permeability thus obtained fall into two distinct groups, dependent on the method of measurement employed. The permeability μ_R (or μ_k), derived from measurements of the resistive losses of the high-frequency circuit, differs from the value μ_L (or μ_n), obtained from measurements of the circuit reactance. This anomaly has been shown by Kittel (1946) to arise from the implicit assumption that μ is a real (rather than a complex) quantity at high frequencies. By confining the measurements to either the real or imaginary component of the circuit impedance, different values of the *apparent* real permeability, μ_R and μ_L , are obtained, neither of which is equal to, or simply related to, the true permeability μ . The measurement of both impedance components of the same circuit element, and the consequent determination of μ , is made difficult by the presence of a high conductivity, equivalent to an imaginary permittivity of large magnitude intimately associated with the permeability. These difficulties may be readily overcome, however, when the material has a low conductivity.

A waveguide impedance method for the measurement of the high-frequency permeability and permittivity of ferromagnetic materials of low conductivity has been developed, and used at wavelengths from 60 cm. to 1.25 cm. The method was primarily developed for the investigation of ferromagnetic compounds, such as magnetite, γ -ferric oxide and the ferrites. It may also be used, however, for the study of ferromagnetic metals, in the form of finely divided powders dispersed in a non-conducting medium. The ferromagnetic compounds are of

particular interest because, due to their low conductivity, the high-frequency magnetic field can penetrate much more deeply than in a metal. It may be possible, therefore, to detect magnetic dispersion and absorption effects due to internal phenomena, which are completely obscured in a metal by the skin-effect. For example, Becker (1938) has given a theory of magnetic dispersion, due to internal eddy-current fields produced by the oscillatory displacements of the domain walls, while Landau and Lifshitz (1935) have predicted resonance phenomena, occurring at the Larmor frequency of the electron spins in the crystalline anisotropy field. Both of these theoretical effects have yet to be verified experimentally.

The permeability μ and permittivity ϵ of a medium, relative to their free-space values, are determined at high frequencies from the electromagnetic wave properties of the medium. These properties may be defined in terms of two parameters, the intrinsic impedance z , relative to free space, and the propagation coefficient γ . The relative intrinsic impedance

$$z = (\mu/\epsilon)^{1/2}, \quad \dots\dots(1)$$

in conjunction with Fresnel's equations, determines the reflection and transmission coefficients of the wave at a boundary of the medium. The propagation coefficient

$$\gamma = i \frac{2\pi}{\lambda} (\mu\epsilon)^{1/2}, \quad \dots\dots(2)$$

where λ is the free-space wavelength, determines the phase velocity (imaginary component of γ) and the attenuation coefficient (real component of γ) of the wave propagated through the medium.

For non-magnetic media $\mu = 1$, and for perfect dielectrics, ϵ is real. For lossy dielectrics ϵ is complex, and may be written

$$\epsilon = \epsilon' - i\epsilon'' = |\epsilon| e^{-i\theta_\epsilon} \quad \left. \begin{array}{l} \text{where} \\ \tan \delta_\epsilon = \epsilon''/\epsilon' \end{array} \right\} \dots\dots(3)$$

is known as the dielectric loss tangent. For ferromagnetic materials at high frequencies it had been found experimentally by the author (1946) that, in accordance with the theory of Arkadiew (1913), μ is also complex, and hence may be written

$$\mu = \mu' - i\mu'' = |\mu| e^{-i\theta_\mu} \quad \left. \begin{array}{l} \text{where} \\ \tan \delta_\mu = \mu''/\mu' \end{array} \right\} \dots\dots(4)$$

may be called the magnetic loss tangent.

It will be seen from equations (1) and (2) that, while a measurement of either of the complex quantities z or γ is adequate to derive ϵ for non-magnetic materials, a measurement of both z and γ is necessary and sufficient for the determination of ϵ and μ for ferromagnetic materials.

§ 2. THEORY OF THE METHOD

Measurements are made on a sample of the ferromagnetic material, filling a section of length d of an empty waveguide. Coaxial waveguides are used at wavelengths of 60–20 cm. and 15–6 cm., and H_{10} rectangular waveguides at 3 cm. and 1.25 cm. The input impedances of the filled section, relative to the characteristic impedance of the empty guide, are measured when the section is terminated in

a short-circuit, and when it is terminated in a closed quarter-wavelength line, which is equivalent to an open-circuit. The relative short-circuit and open-circuit input impedances are given by

$$\left. \begin{aligned} z_{sc} &= z_0 \tanh \gamma_0 d, \\ z_{oc} &= z_0 \coth \gamma_0 d, \end{aligned} \right\} \dots\dots (5)$$

where γ_0 is the propagation coefficient and z_0 is the characteristic impedance of the section relative to that of the empty guide. Hence

$$\left. \begin{aligned} z_0 &= (z_{sc} z_{oc})^{\frac{1}{2}}, \\ \tanh \gamma_0 d &= (z_{sc}/z_{oc})^{\frac{1}{2}}. \end{aligned} \right\} \dots\dots (6)$$

For a coaxial waveguide, operating in its principal mode,

$$\left. \begin{aligned} z_0 &= z = (\mu/\epsilon)^{\frac{1}{2}}, \\ \gamma_0 &= \gamma = i(2\pi/\lambda)(\mu\epsilon)^{\frac{1}{2}}. \end{aligned} \right\} \dots\dots (7)$$

For an H_{10} rectangular waveguide, z_0 and γ_0 are dependent on the cut-off wavelength of the empty guide, λ_c ($=2a$, where a is the guide width), and they are given by

$$\left. \begin{aligned} z_0 &= \mu \left[\frac{1}{\mu\epsilon - (\lambda/\lambda_c)^2} \right]^{\frac{1}{2}}, \\ \gamma_0 &= i(2\pi/\lambda)[\mu\epsilon - (\lambda/\lambda_c)^2]^{\frac{1}{2}}. \end{aligned} \right\} \dots\dots (8)$$

Equation (8) reduces to (7) for λ_c infinite, as in the coaxial case.

Determination of input impedances, z_{sc} and z_{oc} . The relative input impedance is obtained from the measurement of the voltage standing-wave pattern in the empty waveguide, which has negligible attenuation. It is given by

$$z_{\text{input}} = \frac{1/n - i \tan \beta l}{1 - i (\tan \beta l)/n}, \quad \dots\dots (9)$$

where n is the voltage standing-wave ratio (ratio of voltage maximum to minimum), l is the distance of the voltage minimum from the input face of the sample, β ($=2\pi/\lambda_g$) is the phase velocity and λ_g is the measured wavelength, in the empty guide ($=\lambda$, for the coaxial case).

Determination of n . For experimental reasons to be discussed later, n is not measured directly, but is obtained from the width of the voltage curve near the minimum. The voltage distribution on a lossless line may be written

$$V^2 = V_{\min}^2 (\cos^2 \beta x + n^2 \sin^2 \beta x), \quad \dots\dots (10)$$

where V_{\min} is the minimum voltage, and V the voltage at a distance x from the minimum. Rewriting (10) as

$$n^2 = 1 + (V^2/V_{\min}^2 - 1) \operatorname{cosec}^2 \beta x, \quad \dots\dots (11)$$

it is seen that n may be determined from a measurement of x , for a given value of V^2/V_{\min}^2 . In practice V^2/V_{\min}^2 is normally taken as 2, so that (11) simplifies to

$$n^2 = 1 + \operatorname{cosec}^2 \beta x_0, \quad \dots\dots (12)$$

where $2x_0$ is the measured distance between points on each side of the minimum, at which the voltage is $\sqrt{2}$ of its value at the minimum.

Determination of l and λ_g . The distance l from the voltage minimum to the input face of the sample is also measured indirectly. The dial gauge or scale used for recording the position of the high-frequency voltage detector has an arbitrary zero, which must be calibrated. This is done by the measurement of the position b of a voltage minimum, with the guide empty and terminated in a short-circuit (figure 1(a)). This minimum is known theoretically to be $m\lambda_g/2$ from the end of the guide (where m is a positive integer). If, then, c is the position of the voltage minimum when the sample, length d , is introduced into the guide (figure 1(b)),

$$b + m\lambda_g/2 = c + l + d, \quad (13)$$

$$\tan \beta l = \tan \beta(b - c - d).$$

i.e.

The value of $\lambda_g/2$, and hence β , is obtained from the difference between two consecutive values of b .

Computation. The sequence of measurements and computation may be best illustrated by a typical set of results.

Table 1

	d	b	λ_g	c	$b - c - d$	$\tan \beta l$	x_0
Source	Expt.	Expt.	b	Expt.		Eqn. (13)	Expt.
	1.940	8.059	23.422	s.c.	4.279	1.840	0.5362
		-3.652		o.c.	6.252	-0.133	-0.0356
	$\sin \beta x_0$	$1/n$	x_{sc}	x_{oc}	x_0	$\tanh \gamma_0 d$	
Source	Eqn. (12)	Eqn. (9)	Eqn. (9)	Eqn. (6)	Eqn. (6)		
s.c.	0.5388	0.4745	0.6933	0.1610	0.3341	2.075	
o.c.	0.1590	0.1570	$\exp(-i34.25^\circ)$	$\exp(i12.45^\circ)$	$\exp(-i10.9^\circ)$	$\exp(-i23.35^\circ)$	
	$\gamma_0 d$	γ_0	$ \mu $	δ_μ	$ \epsilon $	δ_ϵ	
Source	Chart		Eqn. (7), (4)	Eqn. (7), (4)	Eqn. (7), (3)	Eqn. (7), (3)	
	1.86	0.959	1.20	24.95°	10.7	3.15°	
	$\exp(i75.95^\circ)$	$\exp(i75.95^\circ)$					

s.c. = short circuit; o.c. = open circuit.

All lengths are in tenths of an inch.

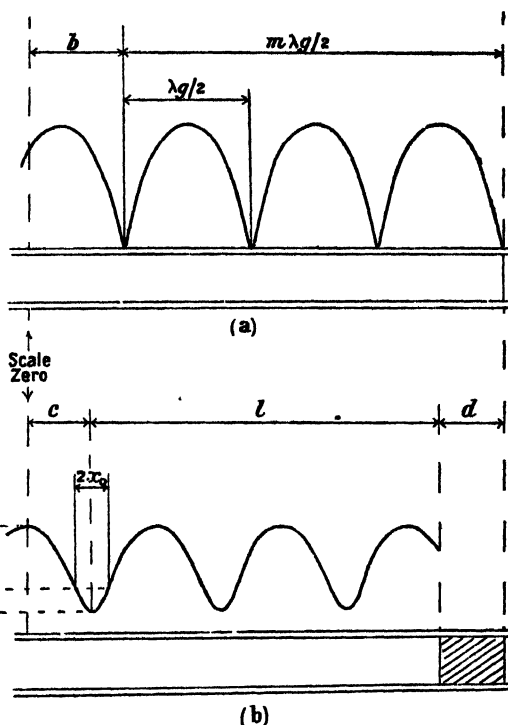


Figure 1. Determination of input impedance.

$\gamma_0 d$ is obtained from the complex hyperbolic tangent, using the *Chart Atlas* compiled by Kennelly (1914). Any ambiguity in the value of $\gamma_0 d$ is removed by the measurement of a second sample of different thickness.

§ 3. DESCRIPTION OF APPARATUS

Waveguide and detector system. The voltage standing-wave is measured by means of a tuned crystal detector, mounted in a sliding carriage, which is moved manually along the length of the waveguide. A short, thin probe from the detector projects into the waveguide through a narrow longitudinal slot, milled parallel to the axis of the guide, and, in the case of the rectangular waveguide, along the centre of its broad side. The high-frequency voltage induced in the probe is rectified by the crystal, and the crystal current is measured by a sensitive galvanometer. Particular care has been taken in the mechanical construction to ensure that the probe moves centrally along the guide, and that its depth of penetration remains constant as the position of the carriage is varied. The tuning of the detector by means of a variable reactance increases its sensitivity and discriminates against the detection of harmonics.

The general arrangement of the complete apparatus used on each of the four wavelength ranges is similar, and it differs only in detail from the schematic diagram of figure 2.

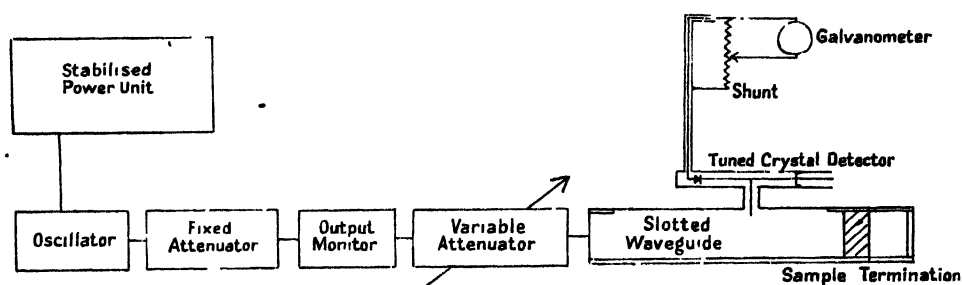


Figure 2. General schematic diagram of apparatus.

60–20 cm. apparatus. A variable-frequency triode oscillator (446 tube with concentric-line tuning) is loosely coupled through a lossy cable to the coaxial waveguide. The oscillator output is monitored at the source by a bolometer bridge circuit. The coaxial waveguide has an outer conductor diameter of 4.45 cm., an inner conductor diameter of 1.35 cm., and a slot length of 110 cm. The inner conductor, which is constructed of light thin-walled tube, is supported at 40 cm. intervals by sets of three thin polystyrene rods, screwed through the outer conductor, at 120° to each other. These spacing rods have been found to have a negligible effect on the wave in the guide. The position of the detector carriage is read on a 0.1 mm. vernier scale. The crystal sensitivity is varied by slightly detuning the detector instead of using a variable attenuator.

15–6 cm. apparatus. At wavelengths near 9 cm. a standard reflex klystron oscillator is used. For other wavelengths in this range a modified oscillator of this type is used (figure 3). The central cylindrical cavity of the oscillator is extended on each side by rectangular waveguide sections, whose length may be varied by short-circuiting pistons, geared to move in synchronism. It is found that

by variation of the transit-time of the electron beam, by adjustment of the electrode voltages, high-order modes of resonance of various wavelengths from 15 cm. to 6 cm. can be induced in this composite cavity. Care must be exercised in the choice of suitable modes, to ensure that the oscillations are stable and monochromatic.

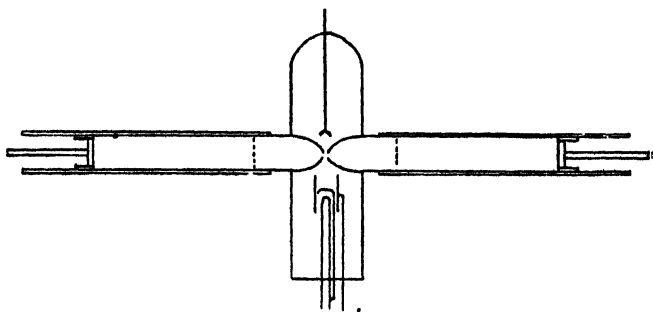


Figure 3. Broad-band reflex klystron.

The klystron oscillator is loosely coupled to a long length of cable feeding the coaxial waveguide, which has an outer conductor diameter of 2.225 cm., inner conductor diameter of 0.775 cm., and a slot length of 16 cm. The inner conductor is supported at the input end by a polythene insulator, and at the load end by the sample and termination. The position of the detector carriage is measured by two 1" micrometer dial gauges mounted on opposite sides of the carriage and recording consecutive 1" traverses. For the longer wavelengths, a single dial gauge and calibrated 1" and 2" sliding steel blocks are used. The oscillator power is monitored by a crystal at the waveguide input, and the sensitivity is varied by slight detuning of the crystal detector.

3 cm. apparatus. A reflex klystron oscillator feeds a $1" \times \frac{1}{2}"$ H_{10} rectangular waveguide system through a short coaxial cable. The sequence of units in the waveguide system is as shown in figure 2. The fixed attenuator, which consists of a flat prism of lossy dielectric material, attached to the broad inner face of the waveguide, has an attenuation of about 6 decibels. The oscillator output is monitored by a fixed crystal detector, or alternatively a resonant cavity wavemeter may be substituted, when required. The variable attenuator consists of a thin circular segment of resistive material, which may be lowered through a longitudinal central slot in the broad face of the guide. The slotted waveguide section is machined from brass plate to obtain the required constructional accuracy. The position of the crystal detector is measured by a micrometer dial gauge attached to the detector carriage, and bearing against a fixed block on the side of the waveguide.

$1\frac{1}{2}$ cm. apparatus. This is similar to the 3 cm. apparatus, except in minor details, with the oscillator feeding directly into a $0.42" \times 0.17"$ H_{10} rectangular waveguide system.

§ 4. EXPERIMENTAL TECHNIQUE

Crystal characteristics. A knowledge of the rectifying characteristics of the crystal detector is essential, particularly in view of the opinion expressed by Collie *et al.* (1946) that "no reliance should be placed on the crystal rectifier obeying an exact law; at best, crystals are unstable, and hard to calibrate". An investigation has therefore been made of the properties of a number of crystals of standard Service type. The bulk of the measurements were made at a wavelength of 9 cm., with the slotted waveguide empty and terminated in a short-circuit.

The crystals were inserted in turn into the detector, and readings of the crystal current I against detector position were taken over a complete quarter-wavelength from minimum to maximum. The voltage distribution on a lossless short-circuited line is from (10), putting $V_{\min} = 0$, $n = \infty$, $nV_{\min} = V_{\max}$,

$$V^2 = V_{\max}^2 \sin^2 \beta x. \quad \dots\dots(14)$$

The crystal current I was, therefore, plotted against $\sin^2 \beta x$. Typical results for a single crystal are shown in figure 4.

The results of the measurements may be summarized as follows. All the crystals investigated have a square-law characteristic (I proportional to V^2) for crystal currents up to a few micro-amperes. At higher currents they deviate in different degrees, often considerably, from the square-law. Mechanical shock and ageing may affect the characteristic, but they do not appear to alter the square law property for low currents. The characteristic is markedly dependent on the galvanometer circuit resistance. The optimum resistance for I to be nearly proportional to V^2 for crystal currents up to $20 \mu\text{a.}$ varies from crystal to crystal, but, is on the average, about 400 ohms.

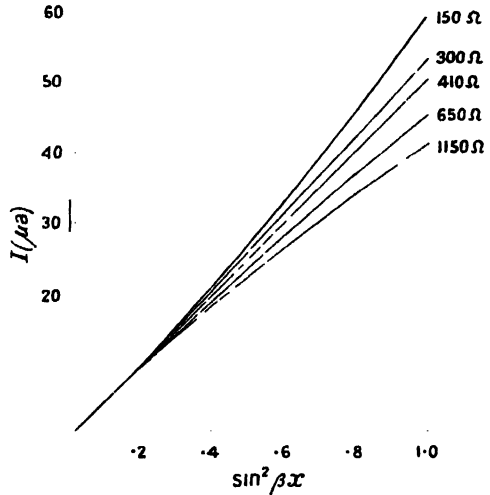


Figure 4. Dependence of crystal law on external resistance.

Impedance measurement. It was, therefore, decided to make all impedance measurements at low crystal currents, using a spot galvanometer with a sensitivity of $1 \mu\text{a.}$ for full-scale deflection, and an internal resistance of 400 ohms. For voltage standing-wave ratios greater than 3, it is not possible to measure the ratio of maximum to minimum current directly with sufficient accuracy, and the indirect method of measuring the width of the voltage curve near the minimum is therefore used (equation (12)).

At the shorter wavelengths with large values of n , the value of $2x_0$ (the width at $2I_{\min}$) is also too small to be measured accurately. It is then necessary to take values of V^2/V_{\min}^2 greater than 2, using (11) to evaluate n . A simple procedure has been employed which avoids the measurement of very low currents. I_{\min} is measured, the current sensitivity of the galvanometer is then reduced by Y ($= V^2/V_{\min}^2$), and the distance $2x$ is measured between points at which the galvanometer current is equal to I_{\min} . A shunt of the type shown in figure 5 is used to avoid changing the resistance of the galvanometer circuit and affecting the crystal characteristic.

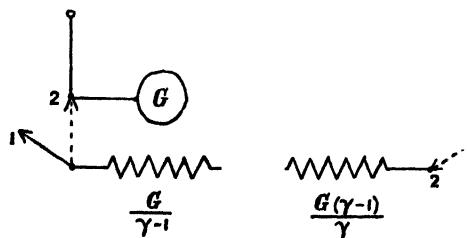


Figure 5 Constant resistance shunt.

Experimental accuracy. A check on the accuracy of each of the waveguide systems has been made, measuring I vs $\sin^2 \beta x$ in an empty short-circuited guide as described above, with a maximum crystal current of $1 \mu\text{a}$. The constancy of the ratio $I/\sin^2 \beta x$ depends not only on the validity of the crystal square law, but also on the overall accuracy of construction and measurement.

Deviations in this ratio might arise from:

- (a) variations in the depth of penetration of the detector probe,
- (b) irregularities in the movement of the carriage,
- (c) rectification of any harmonics present,
- (d) interaction of the detector probe on the waveform, or
- (e) errors in the measurement of current or position.

The results indicate that none of these effects are pronounced, since over the range of currents used in impedance measurements (0.05 to $1 \mu\text{a}$.) the variations of $I/\sin^2 \beta x$ are random, for each of the four waveguide systems, and have the maximum values given in table 2.

Table 2

Wavelength	..	60 cm.	25 cm.	9 cm.	6 cm.	3 cm.	$1\frac{1}{4}$ cm.
Variation	..	$\pm 1\frac{1}{2}\%$	$\pm 2\frac{1}{2}\%$	$\pm 2\frac{1}{2}\%$	$\pm 1\%$	$\pm 1\frac{1}{2}\%$	$\pm 3\%$

No general estimate of error in the determination of ϵ and μ can be given, since it depends on their magnitude and also on the electrical thickness of the sample. Ideally, d should be chosen to give reasonably large values of l and $2x_0$ in both the short-circuit and open-circuit cases. This choice is complicated by the various measured quantities varying in different directions with d , but a suitable compromise can be arrived at by a few preliminary measurements on samples of different thicknesses.

§ 5. PREPARATION AND PROPERTIES OF MIXTURES

Preparation of samples. In studying finely-divided ferromagnetic compounds it is convenient to mix the compound with a lossless, non-magnetic base material to form solid samples for measurement. Paraffin wax has been found very suitable for this purpose. Homogeneous mixtures can be formed by vigorously stirring the compound into molten paraffin wax as it is cooling. The solid mixture is moulded into the required sample shape at room temperature, using a hand-vice for applying pressure. The sample is made slightly oversize, as the material is sufficiently soft to be force-fitted into the waveguide, thereby eliminating experimental errors due to gaps between the sample and the waveguide conductors.

Measurements on paraffin wax. The permittivity of a non-magnetic dielectric can be derived from a measurement of either z_0 or γ_0 , or alternatively from a single measurement of z_{sc} . For a low-loss dielectric, terminated in a short-circuit, equations (9), (5) and (8) simplify and combine to give

$$\left. \begin{aligned} \frac{\tan \phi}{\phi} &= -\frac{\lambda_g \tan 2\pi l / \lambda_g}{2\pi d}, \\ \phi &= \frac{2\pi d [\epsilon - (\lambda / \lambda_c)^2]^{\frac{1}{2}}}{\lambda} \end{aligned} \right\} \dots\dots (15)$$

for the H_{10} rectangular waveguide, and also, putting $\lambda_g = \lambda$, $\lambda_c = \infty$, for the coaxial

waveguide. Equation (15) may be solved using the tables of Jahnke and Emde (1933). Measurements have been made on paraffin wax: ϵ and μ have been calculated from z_{ac} and z_{oc} using the general method; and ϵ has also been calculated from z_{ac} , assuming $\mu = 1$, using (15). Comparative results are given in table 3.

Table 3

	λ in cm.	8.93	3.09	1.236
From z_{ac} and z_{oc} From (15)	ϵ	2.27	2.30	2.28
	μ	0.99	0.98	1.00
	c	2.28	2.29	2.28

Properties of mixtures. Measurements are still in progress on various mixtures of paraffin wax with γ -ferric oxide, magnetic and other ferromagnetic materials, though preliminary reports of some of the results have been published (Birks 1946, 1947). Since it is hoped to give a detailed account of these measurements at a later date, the present discussion will be limited to some typical results obtained with γ -ferric oxide mixtures. These illustrate the applications of the experimental method, and certain other points of general interest.

The relationship between the properties μ , ϵ of the mixtures, and v , the proportion by volume of γ -Fe₂O₃, has been investigated. It is found that, within the limits of experimental error, $|\mu|$ and $|\epsilon|$ vary according to the logarithmic law originally proposed by Lichtenecker (1918) for dielectric mixtures:

$$\log |\mu| = v \log |\mu_a|, \quad \dots (16)$$

$$\log |\epsilon| = v \log |\epsilon_a| + (1-v) \log |\epsilon_0|, \quad \dots (17)$$

where suffix 'a' refers to the (extrapolated) properties of the ferromagnetic powder, and suffix '0' to the properties of the wax ($\mu_0 = 1$). The experimental values of $|\mu|$, obtained for various γ -Fe₂O₃-wax mixtures, at wavelengths from 58.5 cm. to 3.09 cm., are plotted against v in figure 6. The present measurements have been limited to values of v up to 0.41, but Legg and Given (1940), using compressed powdered molybdenum Permalloy, found (16) to be valid for v up to 1.0, and Buchner (1939) and Wul (1946) using rutile mixtures, found (17) to be also tenable up to 100% concentration.

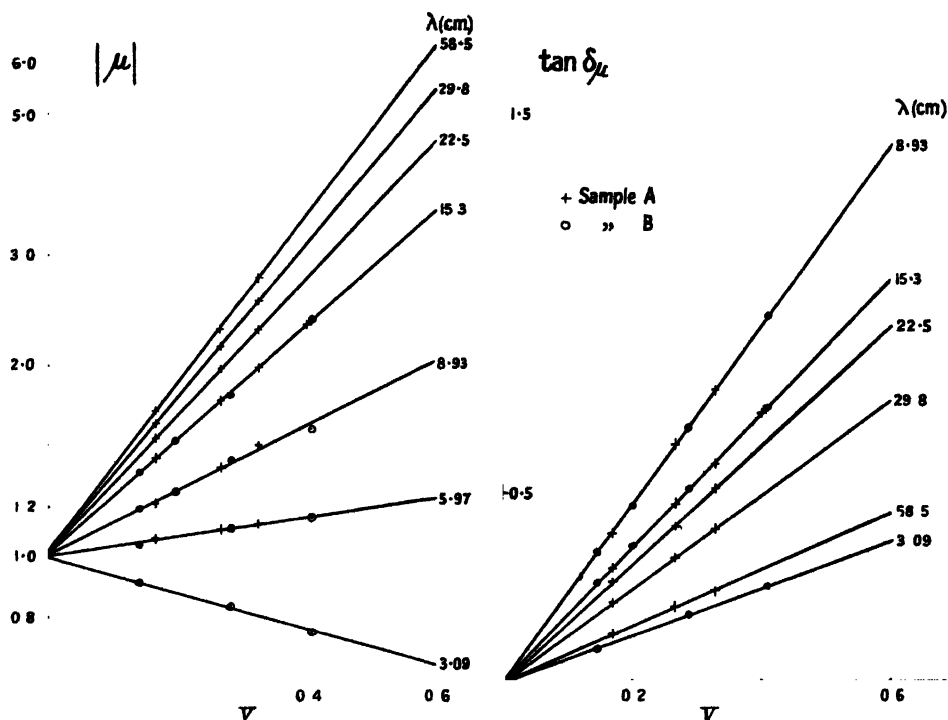
The magnetic and dielectric loss tangents of the mixtures vary linearly with v within the range of the measurements, and may be represented by relations of the form

$$\tan \delta_\mu = v \tan \delta_{\mu_a}, \quad \dots (18)$$

$$\tan \delta_\epsilon = v \tan \delta_{\epsilon_a}. \quad \dots (19)$$

The experimental values of $\tan \delta_\mu$ vs v for different wavelengths are plotted in figure 6. Figure 7 shows the variation with wavelength of the real and imaginary components of μ_a for γ -ferric oxide, derived from these measurements.

Skin-effect. Leigh Page (1971) has shown that, at sufficiently high frequencies, magnetic dispersion and absorption will be exhibited by a medium loaded with


 Figure 6. Magnetic properties of γ - Fe_2O_3 -wax mixtures

conducting ferromagnetic particles, due to skin-effect, even though the ferromagnetic material retains its static permeability. It is, therefore, relevant to enquire whether the results observed with dispersions of ferromagnetic semiconductors at centimetre wavelengths are due wholly or partially to skin-effect.

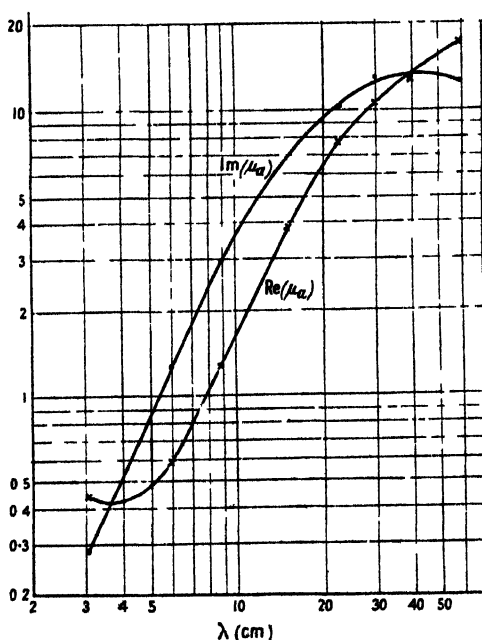
It may be shown from an analysis, similar to that used by Leigh Page, that the ratios of the magnetic and electric moments, p_H and p_E , of a spherical particle, radius r , in a high-frequency electromagnetic field, to the magnetic and electric moments, p_{H_0} and p_{E_0} , in static fields, are given by

$$\frac{p_H}{p_{H_0}} = \frac{\mu + 2}{\mu - 1 + \kappa(\alpha)}, \quad \dots\dots(20)$$

$$\frac{p_E}{p_{E_0}} = \frac{\epsilon + 2}{\epsilon - 1 + \kappa(\alpha)}, \quad \dots\dots(21)$$

where

$$\alpha = \frac{2\pi r}{\lambda} (\epsilon\mu)^{\frac{1}{2}}, \quad \dots\dots(22)$$


 Figure 7. Magnetic spectrum of γ - Fe_2O_3

and
$$\kappa(\alpha) = \frac{\alpha^2 \sin \alpha}{\sin \alpha - \alpha \cos \alpha}, \quad \dots\dots(23)$$

which, for small α , approximates to

$$\kappa(\alpha) = 3 - \alpha^2/5. \quad \dots\dots(23a)$$

For the purpose of the skin-effect analysis only, the properties μ , ϵ , of the ferromagnetic are assumed to retain their static values, thus giving a maximum estimate of any possible skin-effect dispersion, since both μ and ϵ are found experimentally to decrease with increasing frequency. Taking upper limit values for γ -Fe₂O₃ of $\epsilon = \mu = 30$, $r = 0.01$ cm., we obtain from the above equations, at $\lambda = 10$ cm.

$$p_H/p_{H_0} = p_E/p_{E_0} = 1.0002,$$

thus demonstrating that the skin-effect in such a particle is negligible.

An experimental verification of this has been obtained from measurements at three wavelengths on two samples of γ -Fe₂O₃, differing considerably in particle size. If the complex permeabilities observed were due to any significant extent to skin-effect, then they would be critically dependent on the particle radius, r . The measurements on the two samples gave values of μ_a and ϵ_a which are identical within the limits of experimental error, although their radii differed by a factor of about 10. The magnetic properties of the two specimens, referred to as A and B, are distinguished in figure 6.

It may therefore be concluded that the complex permeability observed for γ -ferric oxide and for similar ferromagnetic semiconductors is a property of the material itself, and its further study should therefore provide data on the internal structure of ferromagnetics. It is hoped to discuss the experimental results, in the light of the various theories of ferromagnetic dispersion that have been proposed, in a subsequent paper.

ACKNOWLEDGMENTS

The author wishes to thank Professor P. I. Dee, F.R.S., for the use of the facilities of his laboratory, Professor E. C. Stoner, F.R.S., for his helpful criticism of the preliminary draft of this paper, and the Chief Superintendent, T.R.E., Malvern, for the loan of centimetre-wave valves and other apparatus.

REFERENCES

- ALLANSON, J. T., 1945, *J. Instn. Elect. Engrs.*, **92**, III, 247.
 ARKADIEW, W., 1913, *Phys. Z.*, **14**, 928.
 BECKER, R., 1938, *Phys. Z.*, **39**, 856.
 BIRKS, J. B., 1946, *Nature, Lond.*, **158**, 671.
 BIRKS, J. B., 1947, *Nature, Lond.*, **159**, 775.
 BÜCHNER, A., 1939, *Wiss. Veröff. a.d. Siemens-Werken*, **18**, 84.
 COLLIE, C. H., RITSON, D. M., and HASTED, J. B., 1946, *Trans. Faraday Soc.*, Discussion on Dielectrics.
 GRIFFITHS, J. H. E., 1946, *Nature, Lond.*, **158**, 670.
 JAHNKE, E., and EMDE, F., 1933, *Funktionentafeln*. (Leipzig: B. G. Teuchner.)
 KENNELLY, A. E., 1914, *Chart Atlas of Complex Hyperbolic and Circular Functions*. (Cambridge, Mass.: Harvard University Press.)
 KITTEL, C., 1946, *Phys. Rev.*, **70**, 281.
 LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Sowjet*, **8**, 153.
 LEGG, V. E., and GIVEN, F. J., 1940, *Bell Syst. Tech. J.*, **19**, 385.
 LICHTENECKER, K., 1918, *Phys. Z.*, **19**, 374.
 PAGE, LEIGH, 1941, *Phys. Rev.*, **60**, 675.
 WUL, B., 1946, *J. Phys. U.S.S.R.*, **10**, 95.

On Perfect Optical Instruments

By T. SMITH

National Physical Laboratory, Teddington

MS. received 15 May 1947

ABSTRACT. The angular characteristic function T is used to investigate optical systems of finite focal length in which one extended surface is imaged on another without aberration. The forms of T found earlier to correspond to perfect imagery are shown to be exhaustive.

When perfect imagery is achieved the image surface is similar to the object surface, with the exception that it may be stretched uniformly by arbitrary amounts in each of three mutually perpendicular directions. The length of the optical path between conjugate points of these surfaces is proportional to the distance of either from a fixed plane. In general only one object surface can be imaged without aberration.

In an exceptional group two pairs of conjugate surfaces—parts of centred second-order surfaces—can be perfectly imaged. The two surfaces of the same space have co-incident principal axes, and in each of these directions the product of the two magnifications is equal to the square of the ratio of the refractive indices of the external media. In general the centre of the object is not perfectly imaged. The length of the path from any point of either object surface to its image is invariable. The only elementary example of this group is a single spherical refracting surface.

Another exceptional case is when the conjugate surfaces are plane. It is shown that in addition to being free from all aberrations for a single magnification, a system such as a photographic lens can produce images free from curvature, astigmatism and distortion for all magnifications.

The properties found for both these groups contradict some general conclusions reached by Clerk Maxwell.

Though the existence of perfect instruments is not inconsistent with optical laws, material systems of reflecting or refracting surfaces, whatever their shapes, do not enable this ultimate standard to be fully realized. This shortcoming bears significantly on the methods used in designing optical instruments.

§ 1. INTRODUCTION: MEANING OF "PERFECT"

THE designer of optical instruments seeks to evolve systems which are free from aberrations. In geometrical terms this means that all rays originating at the same object point meet again in a point of the image space. It has been customary to regard as perfect only an instrument in which this condition is satisfied for all points of the object space. Examples of such instruments are a plane mirror and Maxwell's fish's eye (Maxwell 1854). In the latter the refractive index varies from point to point of the object and image spaces, and we shall not consider systems of this kind here. When the external media are homogeneous and isotropic, perfection in the sense just mentioned is only attainable when every image is geometrically similar to the object, and the linear magnification is the inverse ratio of the refractive indices of the two external spaces. The system is therefore afocal (Maxwell 1858, Whittaker 1907). This conclusion follows from the equality of anharmonic ratios for the two spaces, and the equality of all optical paths between given conjugate points. Perfection

in this comprehensive sense is obviously of little interest—it means in practice that the only perfect instruments we can achieve are constructed solely with plane mirrors. Instead we require a definition which can be applied to instruments of finite focal length. The primary aim of the designer of these systems is to image one particular surface on another with the least residual aberrations. This suggests that a perfect optical instrument may be defined as one which deflects all rays emanating from any point of a definite object surface so that they meet again in a single point of an image surface. This is the only kind of perfection of optical significance when the wavelength of the light is considered infinitely small.

The properties of perfect instruments are naturally of special interest, and to investigate them comprehensively it is convenient to use Characteristic Function theory. In this inquiry the function T , the so-called angular characteristic function, will be employed. The variables are the modified direction cosines ξ, η, ζ and ξ', η', ζ' of a ray in the external spaces, and T represents the length of the optical path between the feet of perpendiculars to the ray from the origins of coordinates in these spaces. The adoption of this function implies that afocal systems are not included in the investigation, but this limitation is not important. The choice is advantageous in that the location of the origins is not restricted; for instance, a pair of conjugate points may be chosen. We have first to consider what forms of T are consistent with perfect imagery.

§ 2. DEGREES OF FREEDOM

In the first place we make no assumptions about symmetry. Clearly any function of the modified direction cosines can be expressed as a function of variables u_1, u_2, u_3, \dots where

$$u_p = a_p + b_{p1}\xi + b_{p2}\eta + b_{p3}\zeta - c_{p1}\xi' - c_{p2}\eta' - c_{p3}\zeta',$$

the a 's, b 's, c 's being constants. There cannot be more than six linearly independent variables of this kind, and we first show that if T represents a perfect instrument there cannot be more than three when the origins are suitably chosen. Changes of origin are represented by the addition to T of any u , and as such a term is without effect on the theory which follows it will not be considered.

For brevity denote the differential coefficient of T with respect to u_p by T_p , and the matrices

$$\begin{array}{ccc} / & b_{11} & b_{12} & b_{13} \backslash & & c_{11} & c_{12} & c_{13} \\ & b_{21} & b_{22} & b_{23} & \text{and} & & c_{21} & c_{22} & c_{23} \\ & \dots & \dots & \dots & & & \dots & \dots & \dots \end{array}$$

by b and c respectively. By a fundamental property of the function T the points whose coordinates are given by the equations

$$\begin{aligned} (x \ y \ z) &= (T_1 \ T_2 \ \dots)b \\ (x' \ y' \ z') &= (T_1 \ T_2 \ \dots)c \end{aligned} \quad (1)$$

lie on the incident and emergent rays respectively. It is readily seen that if a unique connection exists between a pair of points on this ray it will relate to these particular points. There will be such a connection if T_1, T_2, \dots can be

expressed in terms of x, y, z or of x', y', z' by solving one of equations (1), for this solution can then be substituted in the other. This requires b and c to be non-singular matrices of the third order, and the number of functions u is therefore limited to three. In general each u includes at least one variable appertaining to each space. From the relations

$$\begin{aligned} \text{or} \quad (x' \ y' \ z') &= (x \ y \ z) b^{-1} c \\ (x \ y \ z) &= (x' \ y' \ z') c^{-1} b \end{aligned}$$

it appears that the conjugate surfaces are similar to one another, though the magnification may vary with direction. In any one direction the magnification is constant, and in this sense there is no distortion. In the absence of the arbitrarily additive u the two origins are conjugate to one another, if either lies on a perfectly imaged surface.

§ 3. T A HOMOGENEOUS FUNCTION

We have not yet taken into account that in a perfect instrument the length of the optical path between an object point and its image is independent of the ray along which it is measured. Denote this path length by P . Then from the meaning of T

$$\begin{aligned} P &= T - (x\xi + y\eta + z\zeta) + (x'\xi' + y'\eta' + z'\zeta') \\ &= T - (u_1 - a_1)T_1 - (u_2 - a_2)T_2 - (u_3 - a_3)T_3. \end{aligned} \quad \dots\dots(2)$$

This length must depend on the coordinates x, y, z (or alternatively on x', y', z') only, whatever the values of u_1, u_2, u_3 may be. This condition must be satisfied either automatically because the right side of (2) is in fact independent of u_1, u_2, u_3 or because u_1, u_2, u_3 can be determined as functions of T_1, T_2, T_3 , and so by (1) P can be expressed as a function of the coordinates of the object or image point.

We shall see later that the second of these alternatives must be rejected. The first implies that T is a homogeneous function of the first order in u_1, u_2, u_3 (Smith 1928). The path length is then

$$P = P_0 + a_1 T_1 + a_2 T_2 + a_3 T_3 \quad \dots\dots(3)$$

or by (1)

$$P - P_0 = (x \ y \ z) b^{-1} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = (x' \ y' \ z') c^{-1} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \dots\dots(4)$$

where P_0 is the length of the path between the two origins when these lie on the conjugate surfaces. The path length for any point of the perfectly imaged surfaces is proportional to the distance of the point from a fixed plane.

The form of the conjugate surfaces depends on the way in which u_1, u_2, u_3 are combined in T . Since T is homogeneous of the first order, T_1, T_2 and T_3 are homogeneous of order zero, and the elimination of the ratios $u_1 : u_2 : u_3$ gives an equation

$$\theta(T_1, T_2, T_3) = 0. \quad \dots\dots(5)$$

The equations of the conjugate surfaces are found by substituting in (5) for T_1, T_2, T_3 from the solutions of equations (1).

For example, if

$$T = u_1 u_2 / u_3,$$

then
$$T_1 = \frac{u_2}{u_3}, \quad T_2 = \frac{u_1}{u_3}, \quad T_3 = -\frac{u_1 u_2}{u_3^2},$$

and therefore

$$T_1 T_2 + T_3 = 0,$$

so that the conjugate surfaces are of the second order and pass through the origins of coordinates.

Another example leading to conjugate surfaces of the second order is

$$T = (au_1^2 + bu_2^2 + cu_3^2 + 2fu_2u_3 + 2gu_3u_1 + 2hu_1u_2)^{1/2},$$

which gives

$$T(T_1 \ T_2 \ T_3) = (u_1 \ u_2 \ u_3)m,$$

where

$$m = \begin{pmatrix} a & h & g \\ h & b & f \\ g & f & c \end{pmatrix}.$$

Since

$$T^2 = (u_1 \ u_2 \ u_3)m$$

the equation connecting T_1, T_2, T_3 is

$$(T_1 \ T_2 \ T_3)m^{-1} \begin{pmatrix} T_1 \\ T_2 \end{pmatrix} = 0$$

and those of the object and image surfaces are of the form

$$(x \ y \ z)M \begin{pmatrix} x \\ y \\ z \end{pmatrix} = 1.$$

The surfaces are therefore of the second order and the origins are at their centres.

§ 4. AXIALLY SYMMETRICAL INSTRUMENTS

Most optical instruments in which fine correction is sought have an axis of rotational symmetry, which will be assumed to coincide with the axes of z and z' . In consequence of this symmetry ξ, η, ξ', η' can only enter T in the combinations $\xi^2 + \eta^2, \xi\xi' + \eta\eta', \xi'^2 + \eta'^2$, and the variables can always be chosen so that

$$T = uf(w), \quad \dots\dots (6)$$

where
$$u = \beta + \gamma\xi - \gamma'\xi', \quad w = \frac{(\alpha\xi - \alpha'\xi')^2 + (\alpha\eta - \alpha'\eta')^2}{u^2}. \quad \dots\dots (7)$$

The length of the optical path is P where

$$P = \beta(f - 2wf'), \quad \dots\dots (8)$$

f' being the differential coefficient of the function f .

In terms of cylindrical coordinates r, z , the conjugate surfaces are obtained by eliminating w between the equations

$$\frac{r^2}{\alpha^2} = \frac{r'^2}{\alpha'^2} = 4wf'^2, \quad \frac{z}{\gamma} = \frac{z'}{\gamma'} = f - 2wf'. \quad \dots\dots (9)$$

The path length and the surface coordinates therefore depend on the single variable w , though this is not true of T . The equations of the surfaces are most simply obtained by expanding z/γ in a series of ascending powers of $r^2/4\alpha^2$, and when the equation of the surface is given in this form an exactly similar process yields f as a power series in w . This is a simple generalization of the normal procedure for constructing T for a single refracting surface (Smith 1927); in that case the conjugate surfaces coincide with the refracting surface, and without loss of generality we can put $\alpha = \alpha' = \gamma = \gamma' = 1$, $\beta = 0$. Systems in which β is zero are of limited interest, since by (8) the path length between conjugate surfaces is zero or at least is invariable over the whole of these surfaces, and this condition is inconsistent with properties desired in many instruments. The group $\beta = 0$ is chiefly important because instrumental design must be based on the properties of single surfaces which are necessarily members of the group.

§ 5. SPHERICAL SURFACES: SYSTEMS WITH TWO PAIRS OF PERFECTLY DEFINED SURFACES

Much interest has been taken in recent years in instruments designed to give good definition over an image surface which is spherical rather than plane. Instruments perfectly corrected for spherical conjugate surfaces are represented by

$$T = kv^3, \quad \dots\dots(10)$$

where
$$v = (\alpha\xi - \alpha'\xi')^2 + (\alpha\eta - \alpha'\eta')^2 + (\beta + \alpha\xi - \alpha'\xi')^2.$$

The conjugate surfaces have radii $k\alpha$ and $k\alpha'$ respectively and are centred at the origins. The path length is given by

$$P - P_0 = z\beta/\alpha = z'\beta/\alpha',$$

in agreement with equations (8) and (9).

Systems with spherical surfaces fall into two groups—in one β vanishes and in the other it does not—with an interesting difference in their properties. This may be illustrated by the simplest possible case, a single refracting surface, in which $\alpha = \alpha' = 1$, $\beta = 0$. As is well known, in addition to the self-conjugate refracting surface, there exists a pair of conjugate spheres on which the imagery is perfect, the magnification being the square of the ratio of the two refractive indices (Huygens 1653). The condition for the existence of these additional surfaces is $\beta = 0$, so that they are necessarily associated with constancy of optical path over the whole of the conjugate surfaces. This peculiarity of having two pairs of aplanatic conjugate surfaces is not limited to a single refracting surface, nor even to systems with spherical conjugate surfaces. For in equation (10) consider the variable

$$v = \{(\alpha + \phi\omega^2)^{\frac{1}{2}}\xi - (\alpha' + \phi\omega'^2)^{\frac{1}{2}}\xi'\}^2 + \{(\alpha + \psi\omega^2)^{\frac{1}{2}}\eta - (\alpha' + \psi\omega'^2)^{\frac{1}{2}}\eta'\}^2 + \{(\alpha + \theta\omega^2)^{\frac{1}{2}}\zeta - (\alpha' + \theta\omega'^2)^{\frac{1}{2}}\zeta'\}^2, \quad \dots\dots(11)$$

where ω and ω' are the reciprocals of the refractive indices μ and μ' . In virtue of the identities

$$\xi^2 + \eta^2 + \zeta^2 = \mu^2, \quad \xi'^2 + \eta'^2 + \zeta'^2 = \mu'^2,$$

it may be seen that v is also given identically by the equation

$$v = \{\omega(\alpha'\mu'^2 + \phi)^{\frac{1}{2}}\xi - \omega'(\alpha\mu^2 + \phi)^{\frac{1}{2}}\xi'\}^2 + \{\omega(\alpha'\mu'^2 + \psi)^{\frac{1}{2}}\eta - \omega'(\alpha\mu^2 + \psi)^{\frac{1}{2}}\eta'\}^2 + \{\omega(\alpha'\mu'^2 + \theta)^{\frac{1}{2}}\zeta - \omega'(\alpha\mu^2 + \theta)^{\frac{1}{2}}\zeta'\}^2. \quad \dots\dots(12)$$

We conclude that in general a system of the $\beta=0$ group which produces perfect imagery over any centred surface of the second order will yield perfect imagery over another concentric second order surface, and that the product of the two magnifications in the direction of any principal axis is equal to μ^2/μ'^2 . But this generalization fails when the system consists of a single refracting surface; for the condition that in one solution the coefficients of ξ and ξ' should be equal, and similarly for the other coordinates, gives

$$\frac{\alpha' - \alpha}{\alpha'^2 - \alpha^2} = \phi - \psi = \theta,$$

so that the refracting surface is spherical.

It would be easy at this stage to draw conclusions which are not warranted. We have shown that one centred conicoid can be perfectly imaged on another without distortion. This apparently means that if P and Q are the extremities of any diameter of the object surface they will have as images points P' and Q' which are the extremities of a diameter of the image surface. If this were correct all rays through the centre C of the one surface would necessarily pass through the centre C' of the other surface, so that these centres would be perfectly imaged conjugate points. The existence of three equally-spaced collinear points P, C, Q which have equally-spaced collinear images P', C', Q' implies that the points at infinity on the incident and emergent rays are conjugate to one another, for these sets of four points form harmonic ranges, and this in turn suggests that the system may be afocal, in which case analysis with the aid of the *T* function is not valid. But a decisive consideration is that we apparently have two separate object points P and Q and two separate image points P' and Q' for which the magnification has the same value. This is possible in an afocal system, but otherwise we must understand that only one of these pairs of points exists. This is not at all surprising. When light is refracted at a single surface we tacitly suppose that we are only dealing with light which has passed through one part of the surface, and, whether or no the surface represented by the algebraic equation is closed, only one refraction occurs. Similarly, when rays are refracted at a spherical surface, the aplanatic surfaces for the magnification μ^2/μ'^2 are not closed; their properties only hold for those parts of the aplanatic spheres, more remote from the point of refraction, determined by the tangent cone from this point to the smaller of the surfaces. It is clear that a similar limitation must hold in the more general cases we have been considering.

The question whether the centres of the perfectly defined surfaces are imaged without aberration has therefore to be considered independently. The conclusion reached is that this perfect imagery is only attained when the surfaces are spherical. In other cases only an incident ray which coincides with one of the principal axes of the object surface will, on emergence, coincide with the corresponding axis of the image surface and so pass through the centre of this surface. When the surfaces are surfaces of revolution a more important union of rays at the centre follows from the indeterminacy of two of the axes.

To show this suppose that

$$v = (\alpha_1\xi - \alpha_1'\xi')^2 + (\alpha_2\eta - \alpha_2'\eta')^2 + (\alpha_3\zeta - \alpha_3'\zeta')^2,$$

so that a point on the object surface is x, y, z where

$$\frac{x}{\alpha_1(\alpha_1\xi - \alpha_1'\xi')} = \frac{y}{\alpha_2(\alpha_2\eta - \alpha_2'\eta')} = \frac{z}{\alpha_3(\alpha_3\zeta - \alpha_3'\zeta')} = \frac{k}{v^{\frac{1}{2}}} \quad \dots (13)$$

This ray passes through the centre of the surface if

$$\frac{x}{\xi} = \frac{y}{\eta} = \frac{z}{\zeta} = \frac{k}{\left\{ \left(\frac{\xi}{\alpha_1} \right)^2 + \left(\frac{\eta}{\alpha_2} \right)^2 + \left(\frac{\zeta}{\alpha_3} \right)^2 \right\}^{\frac{1}{2}}} = \frac{k}{c} \text{ say}$$

Similar equations hold for the image space. Equating the values of x/k etc. given by these equations we have, if the same ray passes through both origins,

$$\frac{\alpha_1\xi - \alpha_1'\xi'}{v^{\frac{1}{2}}} = \frac{\xi}{\alpha_1 s} = \frac{\xi'}{\alpha_1' s'},$$

or

$$v^{\frac{1}{2}} = \alpha_1^2 s - \alpha_1'^2 s',$$

with corresponding equations when α_1, α_1' are replaced by α_2, α_2' and α_3, α_3' . These three simultaneous equations require $\alpha = \alpha_2 = \alpha_3$ and $\alpha_1' = \alpha_2' = \alpha_3'$ if the condition is to be satisfied for rays in all directions. For a ray coincident with a principal axis, say that of x , we have $y = z = \eta = \zeta = 0$, so that η' and ζ' are zero, and hence also y' and z' , i.e. the emergent ray coincides with the axis of x' .

§ 6. SYSTEMS WITH PLANE CONJUGATE SURFACES

This investigation started with the consideration of degrees of freedom, and subsequent developments have depended on the existence of non-singular matrices b and c of the third order. For axially symmetrical systems this has led to the general equations (6) to (9), and we note that these equations are still valid if $\gamma = \gamma' = 0$, when the object and image surfaces become plane. The variables ζ and ζ' have now disappeared from T , and it is clear that in a number of important respects the case is abnormal. The first thing to be noted is that the conjugate surfaces are plane whatever the form of f may be—this is the only case in which the shape of the conjugate surfaces does not determine the characteristic function. Regarded as a function of the single variable w , T is no longer tied to the requirement of being a homogeneous function of the first order.

The normal condition that the path length is proportional to the distance of an object point from a fixed transverse plane obviously does not hold in this case.

Corresponding to different forms of f there will be differences in the aberrations for objects outside the corrected planes. For photographic lenses and other instruments used to image objects at different distances the form of this function is not unimportant. With many of these instruments it is convenient to fit an aperture stop (often an iris diaphragm) effectively in a plane through a principal point, and it is advantageous to choose f so that this point is imaged without aberrations. This condition is satisfied if

$$T = c\{(\alpha - \alpha')^2 - (\alpha\xi - \alpha'\xi')^2 - (\alpha\eta - \alpha'\eta')^2\}^{\frac{1}{2}}. \quad \dots (14)$$

To consider the use of the instrument for magnification m , take new variables

$$h = \frac{(\xi - m\xi')^2 + (\eta - m\eta')^2}{(1 - m)^2}, \quad j = \frac{(\xi - m\xi')(\xi - \xi') + (\eta - m\eta')(\eta - \eta')}{(1 - m)^2}, \\ k = \frac{(\xi - \xi')^2 + (\eta - \eta')^2}{(1 - m)^2},$$

so that h measures the area of the field of view, k the area of the aperture stop, and j is the intermediate variable which occurs in the measure of all unsymmetrical aberrations. Referred to the new axial object and image points the characteristic function is

$$c\{(\alpha - \alpha')(1 - h + 2Mj - M^2k)^{\frac{1}{2}} - M\frac{\alpha}{m}(1 - h + 2mj - m^2k)^{\frac{1}{2}} + M\alpha'(1 - h + 2j - k)^{\frac{1}{2}}\}$$

where

$$M = \frac{m\alpha - \alpha'}{\alpha - \alpha'}.$$

By inspection this function vanishes when $\alpha = 0$, $\alpha' = 0$, or $m = 1$, and when $M = 0$ or $q - (1 - h)k + j^2 = 0$ it reduces to $c\alpha\alpha'(1 - m)^2(1 - h)^{\frac{1}{2}}/m(\alpha - \alpha')$. If it is expanded in a series of ascending powers of q the value is found to be

$$\frac{c\alpha\alpha'(1 - m)^2(1 - h)^{\frac{1}{2}}}{\alpha - \alpha'} \{m^{-1} + \frac{1}{2}Mqpj + \frac{1}{8}Mq^2p^3 \times (\text{quintic in } 1 - h, j) + \dots\},$$

where

$$p^{-1} = (1 - h + Mj)(1 - h + mj)(1 - h + j).$$

The absence of a term in j when $q = 0$ shows that there is no distortion, and the absence of a term in q without j as a factor indicates freedom from curvature and astigmatism with the stop in this special position. Of Seidel aberrations only coma and spherical aberration are present. It follows that, no matter how large the field of view may be, an instrument described by equation (14) is suitable for use over an extended range of magnifications provided the numerical aperture is properly restricted. It will be observed that since k enters only through q , i.e. since k is always associated with $hk - j^2$, all aberrations found when the object surface is changed involve differences between skew and meridional rays.

§7 FAILURE OF OTHER FORMS OF T FOR PERFECT IMAGERY

We have hitherto considered forms of T homogeneous of the first order in the three linear variables u_1, u_2, u_3 , but have found that when T is a function of only two linear variables it need not be homogeneous. The necessity for homogeneity in the general solution of equation (2) is tied up with the number of variables.

To show this it will be sufficient to take simple forms of u_1, u_2, u_3 , each involving one variable from each external space. When T is not homogeneous of the first order the constant path condition shows that rays through a given object point are determined by conditions reducible to the form

$$\alpha_1\xi - \alpha_1'\xi' = \text{constant}, \quad \alpha_2\eta - \alpha_2'\eta' = \text{constant}, \quad \alpha_3\zeta - \alpha_3'\zeta' = \text{constant}.$$

In addition, the two identities

$$\xi^2 + \eta^2 + \zeta^2 = \mu^2, \quad \xi'^2 + \eta'^2 + \zeta'^2 = \mu'^2$$

must be satisfied. These five relations between the six ray variables leave only one degree of freedom, so that absence of aberrations can be secured for a one-dimensional fan of rays, but not for a solid-angular bundle. The only escape from homogeneity therefore occurs when T does not involve all six direction cosines. If only one is missing, say ζ' , the image surface is plane, but the rays from any given object point must make ζ constant, i.e. the object is at infinity;

this case may therefore be treated as a member of the second group, with the two variables ξ and ξ' absent, already considered. We need not consider any other cases for their properties are not of interest. For example, all rotationally symmetrical instruments of finite focal length must have a non-vanishing term in $\xi\xi' + \eta\eta'$.

§ 8. A THEOREM OF CLERK MAXWELL

An elegant proof of the impossibility of a perfect instrument in the wider sense mentioned at the beginning of this paper was given by F. Klein, and has been included in the tract on *The Theory of Optical Instruments* by E. T. Whittaker. The result seems to have been reached first by Clerk Maxwell in a well-known paper (Maxwell 1854), but the fact that his argument is erroneous has apparently escaped notice.* The most important result given by Maxwell is Proposition IX which states:

"It is impossible, by means of any combination of reflections and refractions, to produce a *perfect* image of an object at two different distances, unless the instrument be a telescope and $l = n = \mu_1 \mu_2$, $m = 1$ ". (l and n are the transverse and longitudinal magnifications, and m the angular magnification.)

This proposition is false; it is clearly inconsistent with the result given above that two aberrationless pairs of conjugate surfaces which are in the form of centred conicoids are possible. Maxwell's inference was drawn from an argument intended to show that equality of path cannot be attained for two positions of an object along a ray. It may therefore not be superfluous to note that if x_1, y_1, z_1 and x_1', y_1', z_1' are the coordinates of object and image points when v is defined by (11), and x_2, y_2, z_2 and x_2', y_2', z_2' those when v is put in the form (12), differentiation of T gives

$$\begin{aligned}(x_2 - x_1)v^{\frac{1}{2}} &= k\omega(\alpha'\mu'^2 + \phi)^{\frac{1}{2}}\{\omega(\alpha'\mu'^2 + \phi)^{\frac{1}{2}}\xi - \omega'(\alpha\mu^2 + \phi)^{\frac{1}{2}}\xi'\} \\ &\quad - k(\alpha + \phi\omega^2)^{\frac{1}{2}}\{(\alpha + \phi\omega^2)^{\frac{1}{2}}\xi - (\alpha' + \phi\omega'^2)^{\frac{1}{2}}\xi'\} \\ &= \xi k\omega^2(\alpha'\mu'^2 - \alpha\mu^2),\end{aligned}$$

and

$$\begin{aligned}(x_2' - x_1')v^{\frac{1}{2}} &= k\omega'(\alpha\mu^2 + \phi)^{\frac{1}{2}}\{\omega(\alpha'\mu'^2 + \phi)^{\frac{1}{2}}\xi - \omega'(\alpha\mu^2 + \phi)^{\frac{1}{2}}\xi'\} \\ &\quad - k(\alpha' + \phi\omega'^2)^{\frac{1}{2}}\{(\alpha + \phi\omega^2)^{\frac{1}{2}}\xi - (\alpha' + \phi\omega'^2)^{\frac{1}{2}}\xi'\} \\ &= \xi' k\omega'^2(\alpha'\mu'^2 - \alpha\mu^2),\end{aligned}$$

with similar equations for the other coordinates. Thus

$$\frac{k(\alpha'\mu'^2 - \alpha\mu^2)}{v^{\frac{1}{2}}} = \frac{\mu(x_2 - x_1)}{\omega\xi} = \frac{\mu(y_2 - y_1)}{\omega\eta} = \frac{\mu(z_2 - z_1)}{\omega\xi} = \text{reduction of path in object space,}$$

$$\text{and} \quad \frac{k(\alpha'\mu'^2 - \alpha\mu^2)}{v^{\frac{1}{2}}} = \frac{\mu'(x_2' - x_1')}{\omega'\xi'} = \frac{\mu'(y_2' - y_1')}{\omega'\eta'} = \frac{\mu'(z_2' - z_1')}{\omega'\xi'} = \text{increase of path in image space,}$$

so that the path length from object to image is unaltered by moving the object from position 1 to position 2.

* Since this was written I have been reminded that M. Herzberger has pointed out that Maxwell's conclusion is erroneous. Herzberger's paper should be consulted in connection with the general problem considered here.

Maxwell went on to say: "In this way we might show that we cannot in general have an astigmatic, plane, undistorted image of a plane object. But we can prove that we cannot get perfectly focused images of an object in two positions even at the expense of curvature and distortion", and these conclusions also are wrong. Simultaneous correction of all aberrations over plane conjugate surfaces is consistent with constancy of optical path, and, as we have shown, freedom from curvature, astigmatism, and distortion can be achieved for all positions of the object.

§ 9 THE POWER AND THE UNIT SURFACES

The elementary theory of optical instruments is often based on the existence of unit planes and the focal lengths of a typical system. The focal lengths are immediately derivable from a more fundamental quantity, the power. When we leave the paraxial region the power is associated with particular rays rather than with the system itself, and it is known that the unit surfaces are not in fact plane even where they meet the axis. But both the power and the unit surfaces are valuable concepts whether we are dealing with rays within or without the paraxial region, and a word should be said about them in the case of axially-symmetrical perfect instruments.

In the notation introduced in equations (6) and (7) the power for a typical ray is $u/2\alpha\alpha'f'$, and since the magnification for which the system is corrected is α'/α it follows from the usual equations that the distance from the object surface to the object-space unit surface measured along the ray is $2\mu\alpha(\alpha' - \alpha)f'/u$, and similarly that from the image-space unit surface to the image surface is $2\mu'\alpha'(\alpha - \alpha')f'/u$. If $\alpha = \alpha'$ the unit surfaces are those for which the system is corrected, and so are known. In general the shapes of the unit surfaces will vary with the shapes of the perfectly defined surfaces. When these surfaces are plane u is a constant and f a function of w only. The focal lengths therefore depend only on w , and are fixed when the point of intersection with the object plane is known. The unit surfaces are accordingly spherical. In all other cases the unit surfaces depart from the sphere. For instance, if

$$T = k\{(\alpha\xi - \alpha'\xi')^2 + (\alpha\eta - \alpha'\eta')^2 + (\beta + \gamma\xi - \gamma'\xi')^2\}^{\frac{1}{2}},$$

the corrected surfaces are of the second order, and the power is given by

$$\text{power} = \frac{T}{\alpha\alpha'k^2} = \frac{\alpha\xi - \alpha'\xi'}{\alpha'x} = \frac{\alpha\eta - \alpha'\eta'}{\alpha'y} = \frac{\gamma(\beta + \gamma\xi - \gamma'\xi')}{\alpha\alpha'z}.$$

The argument already used to prove that T must be a homogeneous function then shows that the number of degrees of freedom when the power is constant is less than two unless γ is zero.

§ 10. THE REALIZATION OF PERFECT INSTRUMENTS

We have so far been considering imagery in which freedom from aberrations is consistent with the laws of light propagation, and have reached the conclusion that optical systems with very desirable properties are theoretically possible. The question how these desirable properties are to be achieved in practice immediately arises. Apart from cases which, from our present point of view, are trivial, we know that a plane reflector and a spherical refractor are the only elements which yield aberrationless images, and in all the known systems the path

from object to image is zero, so that the practical need for a real image of a real object is not achieved. Can perfect systems with these practically essential properties be constructed from such unpromising materials, which are all we can use? For systems with a limited number of reflecting or refracting surfaces the answer appears to be definitely no. The best we can expect to do is to approach the theoretically possible standard by balancing aberrations arising at the various refracting surfaces so that the residual errors are, for a given aperture and field of view, small enough to be unimportant in comparison with the size of the images of point sources inseparable from the appreciable magnitude of the wavelength of light. When the numerical aperture or the field of view or both are enlarged this aim will probably necessitate a corresponding elaboration of the lens system. The task facing the lens designer is thus not likely to become easier. If this view of the situation is correct the most promising field for work is the study of aberrations of higher orders than those hitherto fully examined, and particularly, by algebraic analysis, the examination of means by which they can be controlled separately or in combination. Research of this kind, if successful, would at least reduce the waste of energy on schemes that are incapable of yielding a satisfactory outcome, and the advances now being made in the construction of computing machines should make possible the tabulation of any functions which proved to be important for the numerical utilization of the knowledge gained. But whatever may be the stage to which such algebraic work is carried there will always be omitted from the analysis aberrations of still higher orders, and before a design is finished it will be necessary to make sure that they are taken into account. Checks of this kind, if made on paper, will take the form, as now, of tracing selected rays through the proposed system. This, of course, is quite a different matter from using ray-tracing as the normal tool for lens designing, and the methods of ray-tracing adopted may bear little resemblance to those now in common use. It is known that laborious tasks of this kind can be carried out by a suitable fully-automatic machine far more rapidly and with greater accuracy than is ever achieved by the methods employed today; and there is every reason to believe that by the use of such machines notable advances in the performance of optical instruments will be achieved.

§ 11. THE COMBINATION OF OPTICAL SYSTEMS

The view that a scheme for the construction of perfect instruments with useful characteristics cannot be devised may be thought unduly pessimistic, though it will probably not be seriously questioned by experienced designers. The reason for this belief is that there appears to be no means of building up a characteristic function of the type given by equations (6) and (7), with β not zero, by combining elementary systems in which β vanishes. If, with the notation of these equations, we combine two perfect instruments, distinguished by the suffixes 1 and 2 respectively, to form another perfect instrument distinguished by the omission of a suffix, it is readily shown that in general the variables w_1, w_2, w are equal, that the functions f_1, f_2, f are in constant ratios to one another, and that

$$\begin{aligned}\alpha f - \alpha_1 f_1 &= \alpha_1' f_1 - \alpha_2 f_2 = \alpha_2' f_2 - \alpha' f = 0, \\ \gamma f - \gamma_1 f_1 &= \gamma_1' f_1 - \gamma_2 f_2 = \gamma_2' f_2 - \gamma' f = 0, \\ \beta f &= \beta_1 f_1 + \beta_2 f_2.\end{aligned}$$

When T is constructed from the sum of the characteristic functions T_1 and T_2 of the two parts by applying the stationary path conditions, these relations cause the third intermediate direction cosine to disappear from T when two are eliminated and, as the last equation shows, when this elimination is automatic the values $\beta_1 = \beta_2 = 0$ involve $\beta = 0$. But in the typical case encountered in the construction of real instruments the stationary conditions take the form of equations expressing the intermediate ξ and η in terms of the six external direction cosines and also of the third intermediate variable ζ . This variable has to be found from the equation $\xi^2 + \eta^2 + \zeta^2 = \mu^2$ before it can be eliminated from the combined characteristic function. The result is an expression for T which involves all the external direction cosines and is in general less simple than those for the component systems. No plan has been devised which ensures that the T obtained in this way will take the form (6) characteristic of a perfect instrument.

It may be observed that the theory just given must be supplemented where systems with plane conjugate surfaces are combined.

The work described above has been carried out as part of the research programme of the National Physical Laboratory, and this paper is published by permission of the Director of the Laboratory.

REFERENCES

- HERZBERGER, M., 1946, *Ann. N.Y. Acad. Sci.*, **48** (no. 1).
 HUYGENS, CHR., 1653, *Dioptrica*, Prop. XII. (The discovery is often attributed to 'Th. Young.)
 MAXWELL, J. CLERK, 1854, *Camb. and Dubl. Math. Jour.*
 MAXWELL, J. CLERK, 1858, *Quart. J. Pure and Applied Math.*, **11**, and *Scientific Papers*, **I**, 271.
 SMITH, T., 1927, *Trans. Opt. Soc.*, **28**, 225.
 SMITH, T., 1928, *Trans. Opt. Soc.*, **29**, 179.
 WHITTAKER, E. T., 1907, in *Cambridge Tracts in Mathematics and Mathematical Physics*, no. 7.

A New Device for Maintaining Constant Stress in a Rod Undergoing Plastic Extension

BY E. N. DA C. ANDRADE

University College, London

MS. received 11 October 1947

ABSTRACT. A simple mechanical device is described by which constant stress may be maintained during the plastic extension of a wire. The mid-point of the wire maintains its position during extension, so that it may be observed by a fixed microscope.

§ 1. INTRODUCTION

TWO devices have been described in the past which maintain constant stress in a wire extending vertically under a load, the hyperbolic weight (Andrade 1910) and the constant stress bar (Andrade and Chalmers 1932). For work now in progress in our laboratory it is desired to observe microscopically the behaviour of the crystallites in a specimen of polycrystalline metal during extension

under constant stress, with the object of interpreting the β flow and the k flow (Andrade 1910)—which may alternatively be called the transient flow and the permanent flow—in terms of physical processes. For this purpose a device is required which will extend a horizontal specimen by a pull from each end, in such a way that the centre of the specimen maintains its position, while further, as the specimen extends, the pull is automatically diminished in such a way that the force per unit cross-sectional area of the specimen remains constant.

§ 2. GENERAL THEORY

AB and CD are rigid rods, pivoted at B and D so as to be free to turn in the plane of the paper (figure 1). GL and HL are rigid rods pivoted at G and H to AB and CD, and at L to a rod LI', running through a guide M, to which a weight W is attached. PS and QS' are members attached to the cylindrical specimen S'S, so that when, owing to the applied weight W , A and C tend to approach, the specimen is put into tension. It can be seen at once that as the distance AC diminishes and the specimen extends, the force acting longitudinally on the specimen diminishes. It remains to find the conditions that make the stress approximately constant.

Let the length of the specimen be l_0 , $AG = c$, $GB = f$, so that

$$AB/GB = CD/HD = r = 1 + c/f$$

and let $GL = HL = p$ and angle $GLH = 2\theta$.

The horizontal force at G and H is $F = \frac{1}{2}W \tan \theta$ and the axial force on the specimen is $(1/2r)W \tan \theta$.

The length of the specimen is*

$$l = l_0 + 2pr(\sin \theta - \sin \theta_0),$$

where θ_0 is the value of θ when $l = l_0$.

$$l = l_0 \{1 + c(\sin \theta_0 - \sin \theta)\},$$

when

$$c = 2pr/l_0.$$

Putting $\theta = \theta_0 - \alpha$ we have

$$l = l_0 \{1 + c \sin \theta_0 (1 - \cos \alpha) + c \cos \theta_0 \sin \alpha\},$$

which, expanding in terms of α and retaining terms up to α^5 , becomes

$$l = l_0 \{1 + \alpha \cdot c \cos \theta_0 + \frac{1}{2} \alpha^2 \cdot c \sin \theta_0 - \frac{1}{6} \alpha^3 \cdot c \cos \theta_0 - \frac{1}{24} \alpha^4 \cdot c \cos \theta_0 + \frac{1}{120} \alpha^5 \cdot c \cos \theta_0\}.$$

The tension in the specimen $= T = (1/2r)W \tan(\theta_0 - \alpha)$.

Expanding in terms of α , we have

$$\begin{aligned} \tan(\theta_0 - \alpha) &= \frac{\tan \theta_0 - \tan \alpha}{1 + \tan \theta_0 \tan \alpha} \\ &= \tan \theta_0 + \sec^2 \theta_0 \{ -\alpha + \alpha^2 \tan \theta_0 - \alpha^3 (\frac{1}{3} + \tan^2 \theta_0) + \alpha^4 \tan \theta_0 (\frac{2}{3} + \tan^2 \theta_0) \\ &\quad - \alpha^5 (\frac{8}{15} + \tan^2 \theta_0 + \tan^4 \theta_0) \}. \end{aligned}$$

* The motion of A and C is not strictly horizontal, but along the arc of a circle. However, in the actual apparatus the range of angular displacement of A and C will be less than 2° , and throughout the calculation all corrections for the difference between the arc and a horizontal line are negligible.

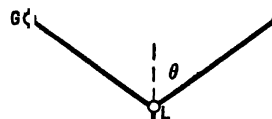


Figure 1.

The required condition for constancy of stress is $Tl\rho = \text{constant}$, where ρ is the density, or, assuming the density constant during extension, which is true to a high degree of approximation, $Tl = \text{constant}$. This means, calling the product p ,

$$p = l/l_0 \tan(\theta_0 - \alpha) = \text{constant},$$

$$\begin{aligned} l/l_0 \tan(\theta_0 - \alpha) = & \tan \theta_0 + \alpha(c \sin \theta_0 - \sec^2 \theta_0) \\ & + \alpha^2 \{ \sec^2 \theta_0 \tan \theta_0 - c \sec \theta_0 (1 - \frac{1}{2} \sin^2 \theta_0) \} \\ & + \alpha^3 \{ c \sin \theta_0 (\frac{1}{2} \sec^3 \theta_0 - \frac{1}{6}) - \frac{1}{3} \sec^2 \theta_0 - \tan^2 \theta_0 \sec^2 \theta_0 \} \\ & + \alpha^4 \sec \theta_0 \{ \tan \theta_0 \sec \theta_0 (\frac{1}{3} + \tan^2 \theta_0) \\ & - \frac{1}{2} c (\frac{1}{3} + \tan^2 \theta_0 + \frac{1}{12} \sin^2 \theta_0) \} \\ & + \alpha^5 \{ c (\frac{1}{120} \sin \theta_0 + \frac{9}{24} \sec \theta_0 \tan \theta_0 + \frac{1}{2} \sec \theta_0 \tan^3 \theta_0) \\ & - \sec^2 \theta_0 (\frac{2}{15} + \tan^2 \theta_0 \sec^2 \theta_0) \} \end{aligned}$$

The conditions for the terms in α and α^2 to vanish are

$$c \sin \theta_0 = \sec^2 \theta_0$$

$$c \sec \theta_0 (1 - \frac{1}{2} \sin^2 \theta_0) = \sec^2 \theta_0 \tan \theta_0,$$

which gives

$$\sin \theta_0 = \sqrt{2/3},$$

$$c = 3\sqrt{3/2}.$$

This makes the coefficients of α^3 , α^4 and α^5 to be -3 , 3.71 and -6 respectively, which, for $\alpha = 10' = 0.17453$ radians, gives for the sum of the three terms -0.0135 in comparison with $\tan \theta_0 = 1.414$, a comparatively small deviation for a range of angles which can be readily made to correspond to a high percentage—say 40% —elongation of the specimen.

§ 3. POINTS AFFECTING DESIGN

The variation of $l/l_0 \tan(\theta_0 - \alpha)$ with α for the case of $c = 3\sqrt{3/2}$ is shown in the lowest curve in figure 2. The two broken horizontal lines represent a variation of 0.1% from the value of $\tan \theta = \sqrt{2}$. The curve representing the product p starts as a horizontal line and then falls at an increasingly rapid rate as α

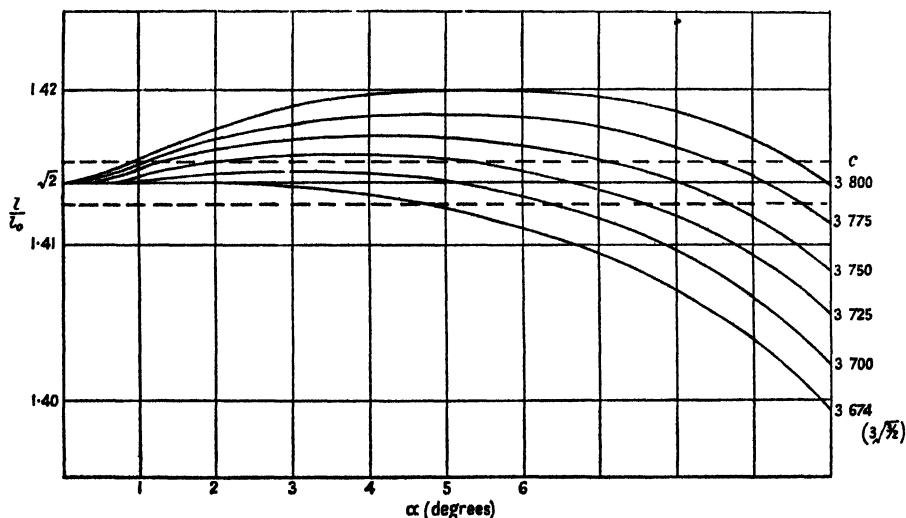


Figure 2.

increases, but up to $\alpha = 4.5^\circ$ it has not changed by 0.1%. Up to $\alpha = 10^\circ$ the variation is 1.0%.

By increasing c slightly the curve is raised in the initial part, as shown, with the result that, if desired, the total variation for a given range of α can be slightly decreased. Curves are shown for values of c up to 3.8. It is evident that for $c = 3.725$ the variation of the product ranges from about +0.12 to -0.12% up to $\alpha = 7.75^\circ$, that is, there is a total variation of 0.24%, while for the same range the $c = 3.674$ curve shows a total variation of about 0.43%. It is clear that the method gives an easy way of maintaining stress constant to a high degree for α up to 8° or so, with a limited range of possibilities of adjusting the slight variations that exist.

It remains to consider what percentage extension of the specimen will correspond to this range of α . The expression

$$(l - l_0)/l = c\{\sin \theta_0(1 - \cos \alpha) + \cos \theta_0 \sin \alpha\}$$

gives the following values:

α	2°	4°	6°	8°	10°
$(l - l_0)/l \times 100$	7.6	15.5	23.8	32.4	41.2

Therefore an extension of about 30%, which is all that is likely to be required, can be obtained with a stress constant to within 0.2%. An extension of 41% involves an overall departure of about 0.44% from constancy of stress.

We are proceeding to the construction of an apparatus along these lines. I am indebted to Mr. A. J. Kennedy for the preparation of figure 2.

REFERENCES

- ANDRADE, E. N. DA C., 1910, *Proc. Roy. Soc. A*, **84**, 1.
ANDRADE, E. N. DA C., and CHALMERS, B., 1932, *Proc. Roy. Soc. A*, **138**, 348.

LETTERS TO THE EDITOR

The Effect of Temperature on the Heat Changes accompanying Magnetization of the Nickel-Silicon Alloy W.5

In a communication to the Society at its meeting in Nottingham on 8 November 1947, F. G. Harrison mentioned the interesting behaviour of the nickel-silicon alloy W.5 (Ni 94.84, Cu 0.1, Fe 0.4, C 0.05, Si 4.0, Mn 0.6), and stated that preliminary experiments had shown that the thermal effects associated with magnetization were very greatly reduced by a rise of temperature of some 25°C . above that of the room.

We have made experiments using the same apparatus modified in such a way that a reasonably steady temperature of 30°C . could be maintained over the specimen. We find that the heat of magnetization given by $\oint H dI$ for a maximum field of 183 oersteds falls from 316 ergs/cm³ per cycle at 20°C . to 212 ergs/cm³ per cycle at 30°C ., while I_{max} falls from 18.3 to 6.1 gauss; obviously, we are dealing with an alloy whose Curie point is near room temperature. At the same time the individual values of the directly measured heat changes were reduced, the maximum cooling observed in any half-cycle falling from 98,460 to 23,300 ergs/cm³.

University College,
Nottingham,
21 January 1948.

L. F. BATES.
J. H. DAVIS.

CORRIGENDA

“The Adiabatic Temperature Changes Accompanying the Magnetization of Cobalt in Low and Moderate Fields”, by L. F. BATES and A. S. EDMONDSON (*Proc. Phys. Soc.*, 1947, **59**, 329).

It is regretted that, by an oversight, the values of $\int I dH$ instead of $\int H dI$ were given in table 8, column 4; columns 4 and 6 and figures 7 and 11*b* should be corrected accordingly. The quantity -3960 should be added to each line, except the first, in column 4 of table 2; the figures in column 5 reduced in the ratio $19/21$ and column 6 recalculated; figures 4 and 11*a* are only slightly affected. The first line in column 3 of table 7 should read 1200 .

“Distribution Coefficients for the Calculation of Colours on the C.I.E. Trichromatic System for Total Radiators at $1500-250-3500^\circ\text{K.}$, and 2360°K. ($C_2 - 14350$)”, by H. G. W. HARDING and R. B. Sisson (*Proc. Phys. Soc.*, 1947, **59**, 814).

Page 824. “ $E_{2750\bar{2}}(0.45\mu)$, 0.1703 ” should read “ 0.1073 ”.

“ $E_{3000\bar{2}}(0.42\mu)$, 1.5102 ” should read “ 1.5012 ”.

“ $E_{3000\bar{2}}(0.48\mu)$, 3.9238 ” should read “ 3.9328 ”.

REVIEWS OF BOOKS

Methods of Measuring Temperature, by EZER GRIFFITHS. Third Edition. Pp. x+223. (London: Charles Griffin and Co., Ltd., 1947.) 20s.

Many of the subjects which receive a chapter in ordinary general text-books are also worthy of whole treatises to themselves, and in some cases they do, in fact, form the subjects of separate treatises. In other cases, however, the recognition of their importance is confined to a relatively small number, and so the demand for full-length treatment is small. Such a subject is that of temperature measurement. Nearly all thermal measurements depend at some point on the measurement of a temperature or temperature difference, and perhaps more measurements have been marred by failure to take adequate precautions here than in any other way, yet the student, hard-pressed for time, generally feels that there is no time for detailed study of it, because it does not loom large in the examination papers.

This book certainly gives the best general survey of the subject at present available, both for measurements at normal temperature and for those at high temperature. It comes from a man of experimental outlook, and consequently contains just those hints that can come only from practice. When it first appeared, it dealt with a subject which was only then emerging into the stage of real accuracy, and now, having been for some time out of print, it has been revised so as to represent the art at its present, much more advanced, stage. In preparing the new edition, the whole text has been re-set, and the book is now much more pleasant to handle than in its old form.

J. H. A.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 60, PART 4

1 April 1948

No. 340

The Electronic Structure and Bond Lengths of Coronene and Pyrene

BY W. E. MOFFITT AND C. A. COULSON*

Physical Chemistry Laboratory, Oxford

* Now at King's College, London

MS. received 1 September 1947

ABSTRACT. Molecular orbital and valence-bond pairing treatments are given for the bond lengths of coronene and pyrene. In both cases the agreement of the molecular orbital treatment with the experimental values is good, and in both cases it is rather better than the agreement of the valence-bond method, though certain characteristic features are common to both treatments and are found experimentally. The Fries rule, which governs the relative importance of the various Kekulé structures in valence-bond theory, is shown to be only partially applicable. A more valid rule is suggested for condensed hydrocarbons generally.

§ 1. INTRODUCTION

THE recent very accurate x-ray work of Robertson and White (1945, 1947) on the dimensions of the polynuclear hydrocarbons coronene and pyrene is of great interest. It discloses a new testing ground for the various theories of the electronic structure of these aromatic systems. In this paper we give the results of two alternative theoretical treatments of these molecules. In the first, using the molecular orbital method of Hund, Mulliken, Hückel and Lennard-Jones, we express the electronic wave function as a product of non-localized one-electron molecular orbitals (called MO), constructed as linear combinations of atomic orbitals (LCAO). In the second (HLSP) method, due to Heitler, London, Slater and Pauling, the total wave function is obtained as a linear combination of many-electron functions or structures, in each of which the electrons are perfectly paired; since these latter calculations would be extremely tedious if followed to the full, we examine an approximation to them.

§ 2. THE MOLECULAR ORBITAL METHOD

The characteristic properties of aromatic systems are attributed to those electrons in orbitals which are antisymmetrical with respect to reflections in the planes of these molecules. We shall consider the behaviour of these (π) unsaturation, or mobile, electrons alone, and make the usual assumption that their interaction with the remaining σ electrons is negligible for our purposes. The LCAO MO procedure has been fully described elsewhere (e.g. Lennard-Jones and Coulson 1939), so that we shall merely quote the results of our calculations for pyrene. The treatment for coronene is of greater interest, however, owing to the high symmetry of the molecule.

The symmetry elements of coronene, like those of benzene, belong to the point group D_{6h} . The 24 $2p_z$ atomic orbitals, one for each carbon atom (figure 1), which we call the $\phi_i (i=1, 2, \dots, 24)$, constitute a reducible representation Γ of this point group. The characters of Γ are found to be

$$\begin{aligned}\chi(E) &= 24, \quad \chi(C_2) = \chi(C_3) = \chi(C_6) \\ &= \chi(C_2'') = 0, \quad \chi(C_2') = -4, \\ \chi(\sigma_h) &= -24, \quad \chi(i) = \chi(S_6) = \chi(S_3) \\ &= \chi(\sigma_v'') = 0, \quad \chi(\sigma_v') = -4.\end{aligned}$$

The notation is fully explained by Wilson (1934). Accordingly, by the theory of group characters, Γ has irreducible components

$$\Gamma = A_{1u} + 3A_{2u} + B_{1g} + 3B_{2g} + 4E_{1g} + 4E_{2u}. \quad \dots (1)$$

Following a method analogous to that used by Bloch (1929) in his discussion of metallic structure, we define the functions

$$\begin{aligned}\kappa(\omega) &= \phi_1 + \omega\phi_2 + \omega^2\phi_3 + \omega^3\phi_4 + \omega^4\phi_5 + \omega^5\phi_6, \\ \lambda(\omega) &= \phi_7 + \omega\phi_8 + \omega^2\phi_9 + \omega^3\phi_{10} + \omega^4\phi_{11} + \omega^5\phi_{12}, \\ \mu(\omega) &= \phi_{13} + \omega\phi_{15} + \omega^2\phi_{17} + \omega^3\phi_{19} + \omega^4\phi_{21} + \omega^5\phi_{23}, \\ \nu(\omega) &= \phi_{14} + \omega\phi_{16} + \omega^2\phi_{18} + \omega^3\phi_{20} + \omega^4\phi_{22} + \omega^5\phi_{24}.\end{aligned} \quad (2)$$

The numbering system of the ϕ_i is as in figure 1, and ω is a number, in general complex, of unit modulus. It may now readily be shown that the required LCAO forms for the molecular orbitals are

$$\begin{aligned}A_{1u}: & \mu(1) - \nu(1), \\ A_{2u}: & k\kappa(1) + l\lambda(1) + m\{\mu(1) + \nu(1)\}, \\ B_{1g}: & \mu(-1) + \nu(-1), \\ B_{2g}: & k'\kappa(-1) + l'\lambda(-1) + m'\{\mu(-1) - \nu(-1)\}, \\ E_{1g}: & k''\kappa(\epsilon) + l''\lambda(\epsilon) + m''\mu(\epsilon) + n''\nu(\epsilon) \\ E_{2u}: & k'''\kappa(\epsilon^2) + l'''\lambda(\epsilon^2) + m'''\mu(\epsilon^2) + n'''\nu(\epsilon^2),\end{aligned} \quad (3)$$

$\epsilon = e^{2\pi i/3} \text{ or } e^{5\pi i/3}$

In these formulae k, l, m, k', \dots, n''' are constants, one set being found from each solution of the secular equation. These sets of coefficients are easily calculated, for, according to (1), the secular equation is a cubic for the levels of symmetry A_{2u} and B_{2g} , it is linear for A_{1u} and B_{1g} , and a biquadratic for the degenerate levels E_{1g} , E_{2u} . The secular equations may be written down when the matrix components of $\mathcal{H} - E$ in the representation (2) are known, where \mathcal{H} is the Hamiltonian and E is the energy, as usual. We have

$$\begin{aligned}(\kappa|\mathcal{H} - E|\kappa) &= 6x\beta + 6(\omega + \omega^5)\beta, \\ (\kappa|\mathcal{H} - E|\lambda) &= (\lambda|\mathcal{H} - E|\mu) = (\lambda|\mathcal{H} - E|\nu) = 6\beta, \\ (\lambda|\mathcal{H} - E|\lambda) &= (\mu|\mathcal{H} - E|\mu) = (\nu|\mathcal{H} - E|\nu) = 6x\beta, \\ (\mu|\mathcal{H} - E|\nu) &= (\nu|\mathcal{H} - E|\mu)^* = 6\omega\beta,\end{aligned}$$

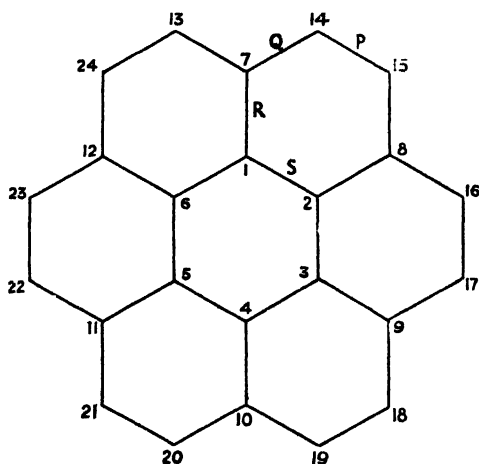


Figure 1. Coronene $C_{24}H_{18}$.

where β is the resonance integral between adjacent $2p_z$ orbitals, E_0 is the atomic energy term (Coulomb term) defined by $E_0 = (\phi_i | \mathcal{H} | \phi_i)$, and $x = (E_0 - E)/\beta$. The secular equations thus reduce to

$$\begin{aligned} {}^1A_{1u}: 0 &= x - 1, \\ {}^1A_{2u}: 0 &= \begin{vmatrix} x+2 & 1 & 0 \\ 1 & x & 2 \\ 0 & 2 & 2x+2 \end{vmatrix} = \Delta(x), \\ B_{1g}: 0 &= x + 1, \\ B_{2g}: 0 &= \begin{vmatrix} x-2 & 1 & 0 \\ 1 & x & 2 \\ 0 & 2 & 2x-2 \end{vmatrix} = -\Delta(-x), \\ {}^1A_{1g}: 0 &= \begin{vmatrix} x+1 & 1 & 0 & 0 \\ 1 & x & 1 & 1 \\ 0 & 1 & x & \epsilon \\ 0 & 1 & \epsilon^* & x \end{vmatrix} = -\frac{1}{2}(x+1)\Delta(-x-1), \\ {}^1A_{2u}: 0 &= \begin{vmatrix} x-1 & 1 & 0 & 0 \\ 1 & x & 1 & 1 \\ 0 & 1 & x & \epsilon^2 \\ 0 & 1 & \epsilon^{*2} & x \end{vmatrix} = \frac{1}{2}(x-1)\Delta(x-1). \end{aligned}$$

Thus the roots are all either ± 1 or simply expressible in terms of the roots of the cubic equation for the levels A_{2u} . The resultant spectrum of energy levels may be represented graphically as in the accompanying diagram (figure 2). In the ground state we fill the orbitals of lowest energy with electrons, two at a time, in accordance with the Pauli exclusion principle. These electrons, whose spins are anti-parallel in pairs, are represented by crosses on our diagram. The final electronic configuration may thus be written

$${}^1A_{1g}: (a_{2u})^2(e_{1g})^4(e_{2u})^4(a_{2u})^2(b_{2g})^2(b_{1g})^2(e_{1g})^4(e_{2u})^4.$$

The explicit LCAO forms for each MO are readily calculated at this stage, since the solutions of the secular equations enable us to determine the values of the sets of coefficients k, l, m, n in (3). That is, each MO may be written in the form

$$\Psi^{(r)} = \sum_i c_i^{(r)} \phi_i \quad (r, i = 1, 2, \dots, 24),$$

where the $c_i^{(r)}$ are known, and satisfy the normalizing condition $\sum_i |c_i^{(r)}|^2 = 1$. The mobile (or π) bond order p_{ij} of a given bond between atoms i and j may now be calculated, for it is the sum of separate contributions from each of the various π electrons. Following Coulson (1938), the calculation

$$p_{ij} = \sum_r p_{ij}^{(r)} = \sum_r \frac{1}{2} (c_i^{(r)} c_j^{(r)*} + c_i^{(r)*} c_j^{(r)})$$

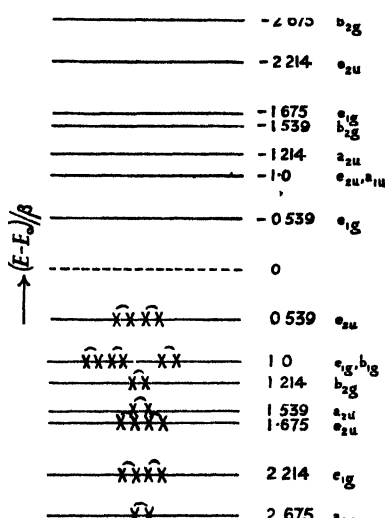


Figure 2 Energies of molecular orbitals in coronene

is readily accomplished. The final values have been tabulated under column 1 for coronene (table 1) and pyrene (table 2). The internuclear distances which these would lead us to predict are given in column 4. Part, but not all, of the information in table 1 has already been given in a preliminary note (Coulson 1944).

§ 3. THE PAIRING METHOD

The HLSP, valence-bond, or pairing, method of treating aromatic molecules consists in selecting that linear combination of the structures ψ_i which gives the minimum energy to the particular molecule under consideration. The procedure is well known and has been applied to many molecules by Pauling (1933) and others. The complete treatment of our molecules along these lines would be an extremely arduous task. For coronene we should have to take more than 200,000 non-ionic structures into account. Therefore most of these structures are neglected, and only a reasonable number are considered to contribute in any marked extent to the ground states of these molecules. It is customary, especially in qualitative discussion, to make the simplification of neglecting all ionic and excited structures containing formal bonds (non-adjacent pairing). Accordingly we confine ourselves to the remaining non-excited, or Kekulé, structures (figure 3), which number 20 and 6 for coronene and pyrene respectively.

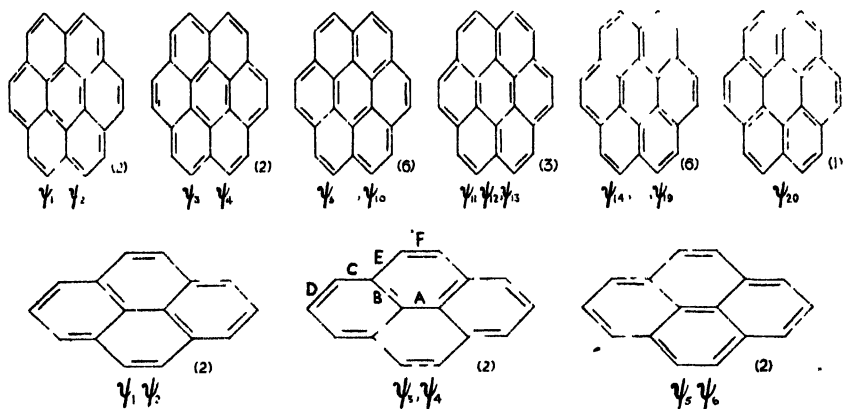


Figure 3. Canonical structures for coronene (top line) and pyrene (bottom line).

By symmetry we may write the two wave functions of the ground states as:

$$\text{coronene } \Psi = a(\psi_1 + \psi_2) + b(\psi_3 + \psi_4) + c(\psi_5 + \dots + \psi_{10}) + d(\psi_{11} + \psi_{12} + \psi_{13}) \\ + e(\psi_{14} + \dots + \psi_{19}) + f\psi_{20},$$

$$\text{pyrene } \Psi = r(\psi_1 + \psi_2) + s(\psi_3 + \psi_4) + t(\psi_5 + \psi_6).$$

The secular equations are of degree 6 and 3 respectively. They may be set up with the help of the symbolic technique devised by Pauling (1933), and they are easily solved. For the two ground states

$$E(\text{coronene}) = Q + 6.64J, \text{ and } a:b:c:d:e:f = 0.50:0.60:0.76:0.75:0.88:1.00,$$

$$E(\text{pyrene}) = Q + 4.70J, \text{ and } r:s:t = 1.31:1.27:1.00.$$

This corresponds to resonance energies $3.64J$ and $2.20J$. Q and J are the Coulomb and exchange integrals as usual.

The bond orders may now be defined in one of two alternative ways. The first is due to Pauling, Brockway and Beach (1935), which we shall abbreviate as (PBB). The second is due to Penney (1937), using a technique devised by Dirac; we shall abbreviate this as (PD). The latter definition of p_{ij} as $p_{ij} = -(\frac{4}{3}) \overline{s_i \cdot s_j}$, where s_i and s_j are the spin matrices of the electrons on atoms i and j , is fundamentally more satisfactory than the earlier one which is merely the weighted mean of double-bond character as revealed by the weights of the various structures. The (PD) bond orders are shown in column 2 of tables 1 and 2; the (PBB) bond orders are shown in column 3. In columns 5 and 6 we give the (PD) and the (PBB) predictions of the bond distances (in Å.), using appropriate curves of order against length in each case. The rows labelled \bar{p} and \bar{r} in the tables give the average bond order and internuclear distances respectively.

Table 1. Coronene

	1	2	3	4	5	6	7
	MO (Coulson)	Pairing (PD) (PBB)		MO (Coulson)	Pairing (PD) (PBB)		Observed (Robertson and White)
P	0.757	0.816	0.742	1.372	1.362	1.362	1.385
Q	0.538	0.316	0.258	1.411	1.450	1.438	1.415
R	0.538	0.629	0.484	1.411	1.387	1.393	1.43
S	0.522	0.328	0.258	1.415	1.446	1.438	1.43
\bar{p}, \bar{r}	0.576	0.481	0.400	1.405	1.419	1.414	1.415

Table 2. Pyrene

	1	2	3	4			
	MO (Coulson)	Pairing (PD) (PBB)		MO (Coulson)	Pairing (PD) (PBB)		Observed (Robertson and White)
A	0.536	0.280	0.231	1.412	1.458	1.445	1.45
B	0.524	0.516	0.384	1.415	1.408	1.410	1.39
C	0.594	0.587	0.500	1.401	1.377	1.390	1.42
D	0.669	0.587	0.500	1.388	1.377	1.390	1.39
E	0.503	0.141	0.116	1.420	1.495	1.484	1.45
F	0.777	0.931	0.884	1.370	1.347	1.348	1.39
\bar{p}, \bar{r}	0.592	0.498	0.421	1.403	1.409	1.412	1.41

Columns 1-3 are bond orders, columns 4-7 are bond lengths.

The last decimal in columns 4, 5 and 6 for the bond lengths (Å.) is unreliable on an absolute scale, but is useful for comparisons.

§ 4. DISCUSSION

Robertson and White (1945, 1947) have determined the dimensions of coronene and pyrene by means of x-ray crystallography, and we have reproduced their values in tables 1 and 2. First, we note that for coronene the agreement between the LCAO MO predictions and the empirical data is very close indeed. The average discrepancy is only 0.01 Å., and in no case is it more than 0.02 Å.—quite within the limits of accuracy of the x-ray determinations. The pairing method gives somewhat poorer results, although the average internuclear distance agrees very well with the measured value. But there are relatively large variations in the lengths of individual bonds, which are apparently not found experimentally. Thus both the x-ray and MO methods gives the bonds R and S a practically equal value, whereas the pairing method expects these to be separated by some

0.05 or 0.06 Å. As Robertson and White have remarked, using only the non-ionic non-excited structures, the empirical data can be reconciled with this representation if the relative weights of the individual structures follow a scheme which is the exact opposite of that which we have calculated; the failure of this set of structures to distinguish between the bonds Q and S is also to be remarked.

It has been suggested (the Fries rule) that those Kekulé structures which show the greatest number of benzenoid rings should have the greatest weight in the superposition diagram. In table 3 we show that this is roughly valid for coronene. Certainly the greatest weight is for ψ_{20} , which has the greatest number (6) of benzenoid rings; and the least weight is for ψ_1 and ψ_2 with the smallest number (1). But apart from these extremes, the weights do not support the rule, which should therefore only be granted a relative validity. A better rule seems to be that those structures which show the greatest number of double bonds in the "exposed" positions (e.g. P of coronene) have the greatest weight. Table 3 shows that there are effectively no exceptions to this in coronene. It seems as if there was a tendency for such outer bonds, in which both ends of the bond are only attached to one other aromatic carbon, to be double; both phenanthrene and pyrene show the same characteristic. For pyrene, shown in table 4, the two types of structure with most exposed bonds (here only 2) have a similar weight, which is greater than that of the remaining structures with only one exposed bond. But the situation regarding the numbers of benzenoid rings is much the same as in coronene.

Table 3. Coronene

Structures	Coefficients	N_b	N_e
ψ_1, ψ_2	0.501	1	3
ψ_3, ψ_4	0.603	4	3
$\psi_5 \dots \psi_{10}$	0.757	3	4
$\psi_{11}, \psi_{12}, \psi_{13}$	0.749	2	4
$\psi_{14} \dots \psi_{19}$	0.879	4	5
ψ_{20}	1.000	6	6

Table 4. Pyrene

Structures	Coefficients	N_b	N_e
ψ_1, ψ_2	1.31	2	2
ψ_3, ψ_4	1.27	3	2
ψ_5, ψ_6	1.00	1	1

N_b =number of benzenoid rings ; N_e =number of exposed bonds.

A similar comparison between theoretical and experimental bond lengths may be made for pyrene as for coronene. We notice, however, that the difference between the MO and x-ray values is larger here than in coronene. This is possibly due to a lack of certainty in the x-ray results, since only nine of the sixteen carbons are "seen" in the diffraction pattern of the crystal. Again the LCAO MO method gives the more likely of the theoretical values. The high double bond character of the exposed bond F is, however, expected by all these methods—in agreement with the chemistry of this molecule.

The relatively poorer values given by the HLSP treatment are readily attributed to the lack of a truly representative set of structures. In conjunction with the findings of Daudel and Pullman (1946) for a rather different set of polynuclear hydrocarbons, we may therefore conclude that in these condensed-ring molecules the contribution of excited structures to the ground states is considerable, since the LCAO MO and the full HLSP calculations using only non-ionic structures generally lead to very similar results for those simpler systems where they can be carried through completely.

REFERENCES

- BLOCH, 1929, *Z. Phys.*, **52**, 555.
COULSON, 1938, *Proc. Roy. Soc. A*, **169**, 413.
COULSON, 1944, *Nature, Lond.*, **154**, 797.
DAUDEL and PULLMAN, 1946, *J. Phys. et Radium*, **7**, 105.
LENNARD-JONES and COULSON, 1939, *Trans. Faraday Soc.*, **35**, 811.
PAULING, 1933, *J. Chem. Phys.*, **1**, 280.
PAULING, BROCKWAY and BEACH, 1935, *J. Amer. Chem. Soc.*, **57**, 2705.
PENNEY, 1937, *Proc. Roy. Soc. A*, **158**, 306.
ROBERTSON and WHITE, 1945, *J. Chem. Soc.*, 607 ; 1947, *Ibid.*, 358.
WILSON, 1934, *Phys. Rev.*, **45**, 706.

The Influence of Surface Films on the Electrical Behaviour of Contacts

BY (MISS) C. C. DILWORTH

H. H. Wills Physical Laboratory, University of Bristol

Communicated by N. F. Mott; MS. received 24 June 1947

ABSTRACT. The variation of current with voltage at an idealized contact between two crystals of a semiconductor is calculated on the assumption that electrons penetrate the surface barrier by tunnel effect. Comparison with experimental curves for silicon carbide powders leads to the conclusion that these crystals are covered by an insulating surface film. The existence of such a film affects the rectifying properties of the crystal when it is in contact with a metal. It is shown that this can account for the discrepancies observed between experimental curves and those deduced from the simple Schottky theory of rectification.

§ 1. INTRODUCTION

THE non-ohmic resistance between a metal and a semiconductor can be due either to the existence of a contact potential between them (Schottky's theory) or to the existence of an insulating layer of different composition (Wilson's theory). In this paper we examine the effect of an insulating layer present in addition to the Schottky layer, the current penetrating the insulating layer by tunnel effect. This model seems correct for silicon carbide with thin quartz layers on the surface. Rectification is in general in the same direction as in the Schottky theory, but the rectifying properties are less good. We consider also the effect of these layers on the contacts between particles of a silicon carbide powder.

§ 2. THE CONTACT BETWEEN TWO SEMICONDUCTORS

The resistance of the contact between two crystals of silicon carbide is of the order of 10^{10} times the resistance of the main bulk of the crystals for potential differences of less than 10^{-2} volt. As the voltage is increased, the resistance falls rapidly, as shown by figure 1, reproduced from the paper by Braun and Busch (1942). We conclude from this and earlier work (Kurtschatow *et al.* 1935) that

this is due to the existence of an insulating film on the surface, and show that it is impossible to explain the magnitude of the effect otherwise.

As a model for calculation, the contact is represented by two spheres, of radius r , in point contact. The area of contact of the surfaces is then zero, but electrons can pass through the air-space in a narrow zone surrounding the point of contact. For a given value of the angle θ between the common axis OX and the radius vector r , the width of the interspace measured in the direction of the common axis is (figure 2):

$$x = 2r(1 - \cos \theta). \quad \dots\dots(1)$$

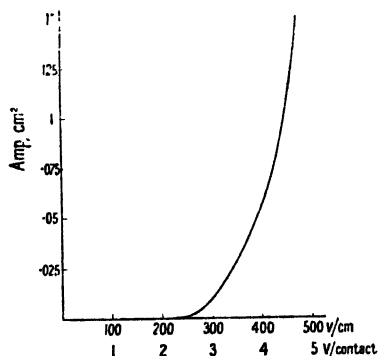


Figure 1. The current-voltage characteristic for silicon carbide crystal contacts.

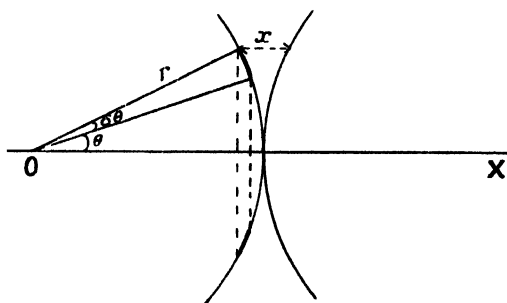


Figure 2. Idealized contact between two spheres.

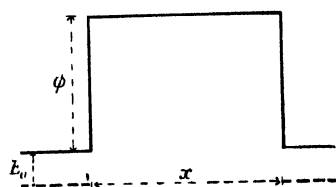


Figure 3. The potential energy of an electron in the interspace.

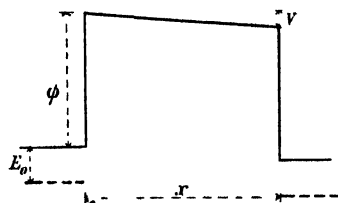


Figure 4. The potential barrier in an external field.

The potential V of an electron in the interspace is represented in figure 3. E_0 is the activation energy of the semiconductor; thus the conductivity depends on temperature according to the formula

$$\sigma = \sigma_0 e^{-E_0/2kT}. \quad \dots\dots(2)$$

$e\phi$ is the work function of the semiconductor; the fraction of incident electrons having sufficient energy to surmount the potential barrier is then

$$f = e^{-e\phi/kT}. \quad \dots\dots(3)$$

Since $e\phi$ is estimated to be of the order of 2-4 e.v., this is negligible at room temperatures ($kT \sim 2.5 \times 10^{-2}$ e.v.). The chief mechanism by which electrons cross the interspace is, therefore, that of "tunnel effect".

On this mechanism the current may be calculated as follows. Under an applied voltage V the potential barrier takes the form shown in figure 4.

The finite probability, P , that an electron with insufficient energy to surmount the barrier should penetrate it is given by

$$P \sim e^{-\alpha x/V}, \quad \dots\dots(4)$$

where

$$\alpha = \frac{4}{3} \eta [(\phi - V)^{3/2} - \phi^{3/2}] \quad \eta = 2m/\hbar,$$

and $\hbar = 2\pi\hbar$ is Planck's constant.

The number of electrons incident on unit area of the barrier per second from left to right is

$$N_1 = \frac{1}{2} N \sqrt{(2kT/\pi m)}. \quad \dots\dots(5)$$

From right to left the number with energy great enough to enter the second conductor is

$$N_2 = \frac{1}{2} N \sqrt{(2kT/\pi m)} e^{-eV/kT}, \quad \dots\dots(6)$$

where N is the number of free electrons in unit volume of the semiconductor. The element of area, normal to the field in the angular range $\theta_0 < \theta < \theta_0 + d\theta$, is

$$dA = 2\pi r^2 \sin \theta_0 d\theta \cos \theta_0. \quad \dots\dots(7)$$

The current crossing this area is

$$\begin{aligned} dI &= e(N_1 - N_2) P dA \\ &= N \{ \sqrt{(kT/2\pi m)} \} (1 - e^{-eV/kT}) e^{-\alpha \cdot 2r(1 - \cos \theta)/V} \cdot 2\pi r^2 \sin \theta_0 \cos \theta_0 d\theta. \dots\dots(8) \end{aligned}$$

Integrating over θ the total current crossing the contact is

$$I = \int_{\theta=0}^{\theta=\frac{1}{2}\pi} dI = N \{ \sqrt{(kT/2\pi m)} \} (1 - e^{-eV/kT}) 2\pi r^2 f(\alpha r/V), \quad \dots\dots(9)$$

where

$$f\left(\frac{\alpha r}{V}\right) = \frac{e^{-\alpha r/V}}{(\alpha r/V)^2} - \frac{1}{(\alpha r/V)^2} + \frac{1}{(\alpha r/V)}.$$

If $r \gg \lambda$, where $\lambda^{-1} = 2\kappa\phi^{1/2}$, then $f(\alpha r/V) \sim V/\alpha r$, and the current is given approximately by

$$I = N \{ \sqrt{(kT/2\pi m)} \} (1 - e^{-eV/kT}) 2\pi r V/\alpha. \quad \dots\dots(10)$$

The integration over the surface from $\theta = 0$ to $\theta = \theta_0$ approximates closely to that over the whole hemisphere for values of θ_0 , such that

$$2r(1 - \cos \theta_0) \geq 10\lambda. \quad \dots\dots(11)$$

This means that the emission of electrons is practically confined to the region $0 < \theta < \theta_0$, and there is an effective area of contact

$$\Delta A = \pi a^2, \quad \dots\dots(12)$$

where

$$a^2 = 2r^2(1 - \cos \theta_0) = 10r\lambda.$$

For such a small area the classical resistance due to the constriction of lines of current flow will be considerable. The contact may be represented by a cylinder of the semiconductor of area πa^2 and length L , where

$$L = (10r\lambda/N) \sqrt{(2\pi m/kT)} \cdot (\alpha/2\pi) (1 - e^{-eV/kT})^{-1}. \quad \dots\dots(13)$$

The total resistance of each contact is then

$$R = (L + \beta a) / \pi a^2 \sigma, \quad \dots (14)$$

where $\beta \sim 0.82$ (J. H. Jeans, *Electricity and Magnetism*).

The specific resistance ρ of a mass of crystals of mean diameter $2r_0$, each in contact at a point of average curvature, will be given by

$$\rho = 2r_0 R = (2r_0 / \pi a^2 \sigma) (L + \beta a). \quad \dots (15)$$

Curve (a) in figure 6 is drawn in the case $r = r_0$, $\phi = 4.5$ volt and $\sigma = 0.5 \text{ ohm}^{-1} \text{ cm}^{-1}$. For $V < kT/e$ this gives a resistance per contact of the order 10^{-2} ohm , so that it is clear that ρ is of the correct order of magnitude only for $\rho : \rho_0 \sim 10^{-8}$, which is not the case. Also the form of equation (10) shows a much slower increase of current with voltage than is actually observed. It therefore appears necessary, in order to obtain a sufficiently high resistance, to assume that the surfaces are separated by an insulating film. The contact is now represented by figure 5.

The thickness of the barrier

$$x = 2r(1 - \cos \theta) + \delta_0, \quad \dots (16)$$

where $\frac{1}{2}\delta_0$ is the thickness of the film on each surface. On substituting this value for x in equation (8), the current is given by

$$I = (\sigma/N) \sqrt{(2\pi m/kT)} (1 - e^{-eV/kT}) 2\pi r^2 f(\alpha r/V) e^{-\alpha \delta_0}. \quad \dots (17)$$

In the approximate form this is

$$I = (\sigma/N) \sqrt{(2\pi m/kT)} (1 - e^{-eV/kT}) (2\pi r^2/V\alpha) e^{-\alpha \delta_0}, \quad \dots (18)$$

and the specific resistance is

$$\rho = (2r_0/\sigma) (L' + \beta a) / \pi a^2, \quad \dots (19)$$

where

$$L' = (10r\lambda/N) \sqrt{(kT/2\pi m)} (1 - e^{-eV/kT})^{-1} e^{\alpha \delta_0}. \quad \dots (20)$$

Curves (b), (c) and (d) in figure 6 correspond to values of $\delta_0 = 1, 2$ and 10 \AA , respectively. The resistance for $V < kT/e$ is of the right order for $\delta_0 \sim 10 \text{ \AA}$. Figure 7 shows how the shape of the theoretical curve compares with experiment. The agreement is not good. Busch's results (Braun and Busch 1942) for rounded grains agree well, but Mitchell's results (Mitchell and Sillars 1947) differ widely from the theoretical curve. Since, however, he finds some rectification even between crystals of the same type, his contacts cannot be of the simple symmetrical form considered and agreement is not to be expected. He does not observe the inflection of the curve at $V \sim kT/e$ due to the factor $(1 - e^{-eV/kT})$ and ascribes this to "leakage" through the barrier possibly due to gaps in the oxide layer. It seems more likely to be due to the rectifying properties already noted.

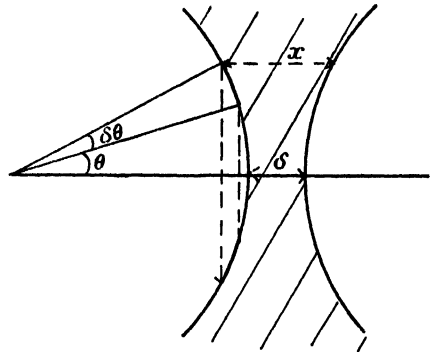


Figure 5. Contact between two semi-conductors through a surface film.

Evidence on the temperature-dependence of the resistance of the barrier is inconclusive. According to tunnel effect mechanism it should be the same as that of the specific resistance of the crystals.

Electron diffraction experiments indicate that there probably exists a surface film of the thickness required.

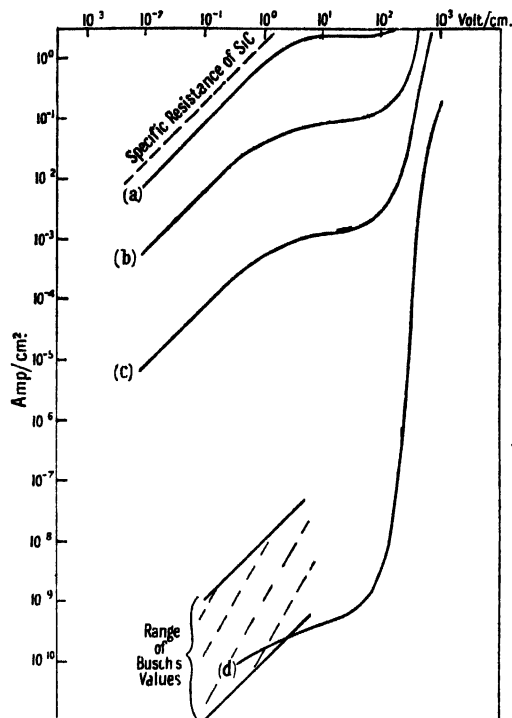


Figure 6. Silicon carbide powder. Theoretical curves.

50 contacts/cm. Work function $\phi = 4.5$ v.

Film thickness

(a) $\delta = 0$. (c) $\delta = 2$ Å.

(b) $\delta = 1$ Å. (d) $\delta = 10$ Å.

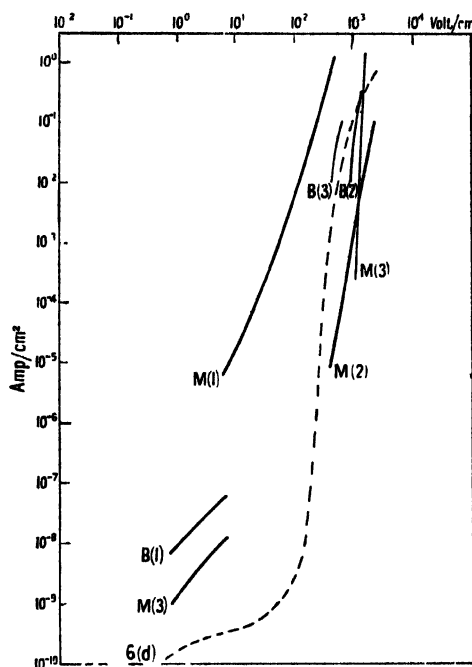


Figure 7. Silicon carbide powder. Experimental curves.

Braun and Busch :

B (1) Region of validity of Ohm's law.

B (2) Rounded grains.

B (3) Angular grains.

Mitchell and Sillars :

M (1) Bonded disc ; particle diam. 0.005 cm.

M (2) " " " " 0.015 cm.

M (3) " " " " 0.025 cm.

6 (d) $\delta = 10$ Å. ; $\phi = 4.5$ v. (50 contacts/cm.).

We conclude that the theory that conduction at these contacts takes place by tunnel effect through a surface barrier accounts for the magnitude, and approximately for the variation with voltage, of the resistance of the contact. The problem to be considered next is that of the nature of a contact between such a crystal and a metal.

§ 3. THE CONTACT BETWEEN A SEMICONDUCTOR AND A METAL

In this section the conduction is always taken to be due to the presence of free electrons. The results can be applied to defect semiconductors, in which positive holes carry the current, by a change in sign of the electronic charge.

The resistance of the contact between a metal and a semiconductor depends not only on the magnitude but also on the direction of the applied voltage. The early theory of rectification, proposed by A. H. Wilson (1932), in which he assumed the existence of a surface barrier penetrated by tunnel effect, gave the wrong sign to the rectification. The more recent theories of Schottky (1939), Mott (1939) and Bethe (1942), which give the right sign, assume that no tunnel effect occurs. These theories give rectification in the observed direction. It will be shown here that a combination of the two theories—a Schottky barrier with an insulating barrier in addition, penetrated by tunnel effect—still gives rectification in the right direction.

Wilson (1937) took a barrier of the form shown in figure 8. He assumed exchange of electrons to take place entirely by tunnel effect, and obtained an expression for the current

$$i = -eN\sqrt{(kT/2\pi m)}(e^{eV/kT} - 1)e^{-\beta\delta}, \quad \dots\dots(21)$$

$$\beta = 2\eta\phi^{1/2}.$$

V is positive when the metal is at the higher potential. According to equation (21) the current is greater then, than when an equal, negative voltage is applied. This is contrary to experience.

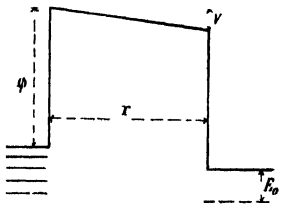


Figure 8. Wilson's potential barrier between a metal and a semiconductor.

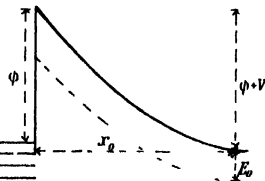


Figure 9. Schottky's space-charge barrier.

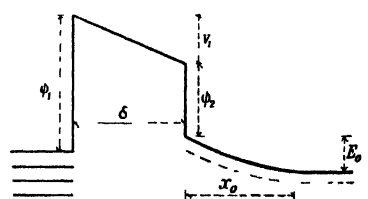


Figure 10. Surface film and space-charge barrier.

Schottky (1939) assumed that the only potential barrier at the contact is that set up within the semiconductor by space-charge, when electrons are lost to the metal by the impurity centres nearest the surface. This type of barrier is shown in figure 9. $e\phi$ is the height of the conduction band of the semiconductor above the Fermi level in the metal, and is given by the difference $e(\phi_1 - \phi_2)$ of the work functions of the metal and the semiconductor respectively. In general $e\phi \gg E_0$, so that for simplicity it may be considered that all the impurity centres at distances up to x_0 from the surface are ionized. Then by Poisson's equation, $d^2V/dx^2 = 4\pi\rho/\kappa$,

$$V = (2\pi Ne/\kappa)(2xx_0 - x^2), \quad \dots\dots(22)$$

where the total potential drop

$$V_0 = \phi + V = 2\pi Nex_0^2/\kappa. \quad \dots\dots(23)$$

In order to calculate the current, either "diode theory" (Bethe 1942) or diffusion theory can be used. In the former it is assumed that the barrier is thin compared to the mean free path of the electrons. Then on the average an electron will not suffer collision in passing over the barrier, and the current will be given by

$$i = (\sigma/N)\sqrt{(2\pi m/kT)}(1 - e^{-eV/kT})e^{-e\phi/kT}. \quad \dots\dots(24)$$

When the barrier thickness is very much greater than the mean free path the diffusion theory must be applied. The current is then obtained from the equation

$$i = e \left\{ -\frac{n(x)}{kT} \cdot \frac{dV}{dx} - \frac{dn}{dx} \right\}, \quad \dots\dots(25)$$

where the mobility $v = eD/kT$, and D is the diffusion coefficient. Using the integrating factor $e^{eV/kT}$, this gives for the current,

$$i \int_0^{x_0} e^{eV/kT} dx = eD \{A - n(x)e^{-eV/kT}\}, \quad \dots\dots(26)$$

where A is a constant.

When $i = 0$, $V(x_0) = \phi$ and $n(x_0) = N$, so that

$$A = Ne^{-e\phi/kT}.$$

When $i \neq 0$, $V(x_0) = \phi + V$, therefore

$$\begin{aligned} i \int_0^{x_0} e^{eV/kT} dx &= eND e^{-e\phi/kT} (1 - e^{-eV/kT}) \\ &= kT \sigma e^{-e\phi/kT} (1 - e^{-eV/kT})/e. \end{aligned} \quad \dots\dots(27)$$

The integral on the left hand side depends chiefly on the field at the surface and varies much more slowly with V and T than the terms on the right. In fact, for $V \ll \phi$, when the barrier is altered little by the applied voltage, equations (24) and (27) are indistinguishable. With the usual assumption that $e\phi \gg E_0$, V has the form given by equation (23), i.e.

$$\int_0^{x_0} e^{eV/kT} dx = \int_0^{x_0} e^{2\pi N e^2 (x^2 - 2xx_0)/\kappa kT} dx.$$

The value of the integrand is a maximum at $x = 0$, and decreases rapidly as x increases. When $x_0^2 > 1/4\pi Ne^2$, the integral has the approximate form

$$\int_0^{x_0} e^{eV/kT} dx = \int_0^{x_0} e^{-4\pi N e^2 x x_0/\kappa kT} dx.$$

Since practically the whole potential drop takes place in the blocking layer, values of $x > x_0$ are equivalent to $x = \infty$. The integral can then be evaluated and gives

$$\int_0^\infty e^{-4\pi N e^2 x x_0/\kappa kT} dx = \kappa kT / 4\pi N e^2 x_0. \quad \dots\dots(28)$$

The current is then

$$i = \sigma \sqrt{\{8\pi N e \kappa (\phi + V)\} (1 - e^{-eV/kT}) e^{-e\phi/kT}}. \quad \dots\dots(29)$$

Both equations (24) and (27) show rectification in the direction commonly observed.

The contact between a metal and a crystal of the type considered in § 1 must take place through a surface film. Both types of barrier will then be present, and the contact may be represented by figure 10.

The total potential drop across the barrier falls into two parts, V_1 from $x = -\delta$ to $x = 0$, and V_2 from $x = 0$ to $x = x_0$. Since there is no change on the surface, the field at $x = 0$ must be the constant field in the insulator and therefore

$$\begin{aligned} V_1 &= F_0 \delta_0 = 4\pi N e x_0 / \kappa, \\ V_2 &= \int_0^{x_0} F dx = 2\pi N e^2 x_0^2 / \kappa, \end{aligned} \quad \dots\dots(30)$$

where

$$V_1 + V_2 = \phi + V = \phi_1 - \phi_2 + V.$$

ϕ_1 and ϕ_2 may be assumed to be so large that transfer of electrons takes place entirely by tunnel effect. For small voltages, i.e. $|V| < \phi$, the current calculated on the diode theory is

$$i = (\sigma/N)\sqrt{(2\pi m/kT)}e^{-2\kappa\phi_1^{1/2}\delta} [e^{-e(\phi-V_1)/kT} - e^{-eV_1/kT}]. \quad \dots\dots (31)$$

On diffusion theory it is

$$i = \{ss'/(s+s')\} [e^{-e(\phi-V_1)/kT} - e^{-eV_1/kT}]e^{-2\kappa\phi_1^{1/2}\delta}, \quad (32)$$

where

$$s = \sigma\sqrt{\{4\pi Ne^2\kappa(\phi+V)\}},$$

$$s' = \sigma/N\sqrt{(2\pi)}.$$

Equations (31) or (32) reduce to (24) or (27) when $V_1 = 0$, i.e. $\delta = 0$, and to (21) when $V_2 = 0$, i.e. $\phi_1 \sim \phi_2$, and, therefore, $x_0 = 0$. In the intermediate case, rectification may be in either direction, but will not be so pronounced, and the current in the direction of high resistance will not saturate at $V = kT$. This is illustrated by figure 11.

The equations (21), (31) and (32) are derived on the assumption that an appreciable current is obtained under a small potential difference. If either ϕ or δ is large this is not so. Then a voltage must be applied to the contact, which is comparable to the contact potential difference before an observable current is obtained. The current from a metal, through a surface barrier in strong fields, is given by the Nordheim-Fowler formula (1928)

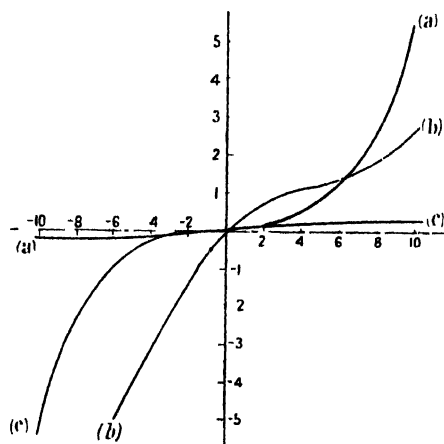


Figure 11. Alternative types of rectifying

(a) $i = \text{const. } (e^{-eV/kT} - 1)$.

(b) $i = \text{const. } \{e^{-e(\phi-V_1)/kT} - e^{-eV_1/kT}\}$
($\delta/\lambda = 1$; $x_0/\lambda = 5$).

(c) $i = \text{const. } (1 - e^{-eV/kT})$.

$$i = \frac{\mu^{1/2}}{2\pi\hbar(\phi_1 + \mu)\phi^{1/2}} \left\{ \delta - 1 \right\} e^{-4\kappa\phi_1^{1/2}\delta/3V} \quad \dots\dots (33)$$

From the semiconductor it is

$$i = (\sigma/N)\sqrt{(2\pi m/kT)}e^{-4\kappa\phi_1^{1/2}\delta/3V}, \quad \dots\dots (34)$$

When $\phi_2 < \phi_1$, the latter current will be the greater. Wilson's theory then gives the right direction of rectification.

Since the space-charge is eliminated when $V < -\phi$, the resistance of the contact vanishes in Bethe's theory and there remains only the resistance of the semiconductor. For $V \geq kT/e$, equation (31) shows that the current reaches the value

$$i = (\sigma/N)\sqrt{(2\pi m/kT)}, \quad \dots\dots (35)$$

which is independent of the applied voltage. In strong fields, however, the barrier is distorted by the image force, so that for $V \sim \phi$

$$i = (\sigma/N) \sqrt{(2\pi m/kT)} e^{-e^{3/2} F^{1/2} / \kappa^{1/2} kT}. \quad \dots\dots(36)$$

A voltage such that $V_1 = -\phi$, applied to the surface film contact, removes the space-charge barrier, so that the current is limited only by the barrier of the surface film. For $-\phi_1 \ll V < -\phi$,

$$i = (\sigma/N) \sqrt{(2\pi m/kT)} e^{-2\kappa\phi_1^{1/2} \delta (1 - (V + \phi)/4\phi_1)}, \quad \dots\dots(37)$$

and for $V \ll -\phi_1$

$$i = (\sigma/N) \sqrt{(2\pi m/kT)} e^{-4\kappa\phi_1^{1/2} \delta / 3 (V - \phi)} \quad \dots\dots(38)$$

when $V_1 \geq \phi_1$ electrons are drawn from below the Fermi level of the metal. The number of electrons crossing against the field is negligible, and the current is therefore given by

$$i = e \int_{\mu - (\phi + V_1)} N(w) P(w) dw, \quad \dots\dots(39)$$

where

$$P(w) \sim e^{-\{2\kappa(r-w)^{1/2} (1 - v/4(c-w))\}},$$

$$N(w) \sim (4\pi me/h^3)(\mu - w).$$

Integrating, this gives

$$i = \frac{4\pi me^2}{h^3} \cdot \frac{e^{-\alpha(1 - V_1/4\phi_1)}}{\alpha^2(1 + V_1/4\phi_1)^2} (2\phi_1)^2 [1 - e^{-\alpha(1 + V_1/4\phi_1)(V_1 - \phi)/2\phi_1 \{1 + \alpha(1 + V_1/4\phi_1)\}}], \quad \dots\dots(40)$$

where

$$\alpha = 2\kappa\phi_1^{1/2} \delta.$$

When $V_1 \geq \phi_1$, equation (33) may be applied, with V replaced by V_1 , i.e.

$$i = \frac{e}{2\pi h} \cdot \frac{\mu^{1/2}}{(\phi_1 + \mu)\phi_1^{1/2}} \left(\frac{V_1}{\delta}\right)^2 e^{-4\kappa\phi_1^{1/2} \delta / 3 V_1}, \quad \dots\dots(41)$$

where, from equation (30),

$$V_1/V = 2x_0\delta/\lambda^2.$$

This set of expressions, together with (31), covers the whole range of voltage for the surface film contact. Figures 12 (a), (b) and (c) have been drawn for suitable values of the constants. The number of unknown parameters, N , ϕ_1 , ϕ_2 and δ , is too great to allow a quantitative comparison with experiment, but certain conclusions can be drawn from the shape of the curves obtained.

Figures 13, 14 and 15 are examples of experimental curves obtained for three different kinds of crystals. They are compared with the theoretical curves to which they most closely correspond. If the values assumed for N , ϕ_1 and ϕ_2 were of the right order of magnitude, it would appear that any film (if it exists) on germanium crystals is less than 5 Å. thick and has no influence on the rectification. On the silicon crystals it is between 5 and 10 Å. and on silicon carbide between 10 and 20 Å. thick. The effect of increasing film thickness, for $|V| < \phi$, is to decrease the degree of rectification and to prevent the saturation of current in the direction of high resistance.

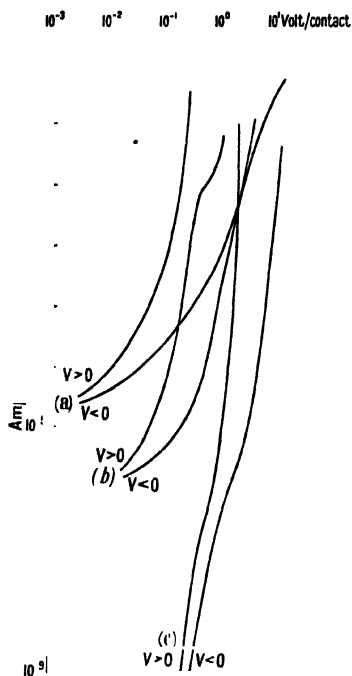


Figure 12. The current-voltage characteristic of the contact between a metal and a semiconductor, for three thicknesses of the surface film.

$\phi_1 = 2.0$ volt. (a) $\delta/\lambda = 0.5$.
 $\phi_2 = 1.5$ volt. (b) $\delta/\lambda = 1.0$.
 $N = 10^{19} \text{ cm}^{-3}$. (c) $\delta/\lambda = 2.0$.

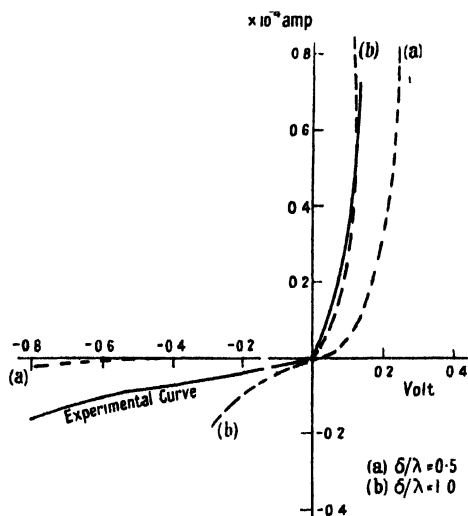


Figure 14. Curve showing rectification in the region $V \sim \phi$ at the contact between silicon and a metal. Experimental data supplied by British Thomson-Houston Co. Ltd.

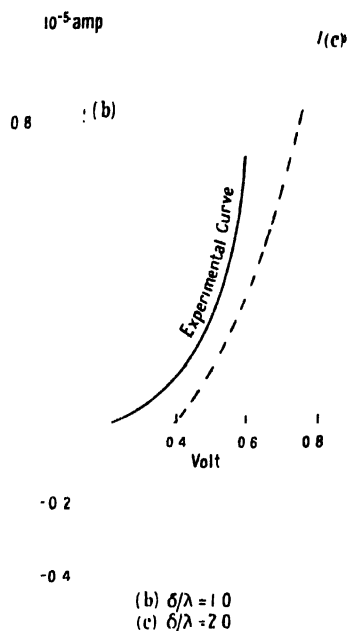


Figure 13. Curve showing rectification in the region $V \sim \phi$ at the contact between silicon carbide and a metal. Experimental data reproduced from a report by L. W. J. Mitchell and R. W. Sillars, by permission of Metropolitan-Vickers Electrical Co. Ltd.

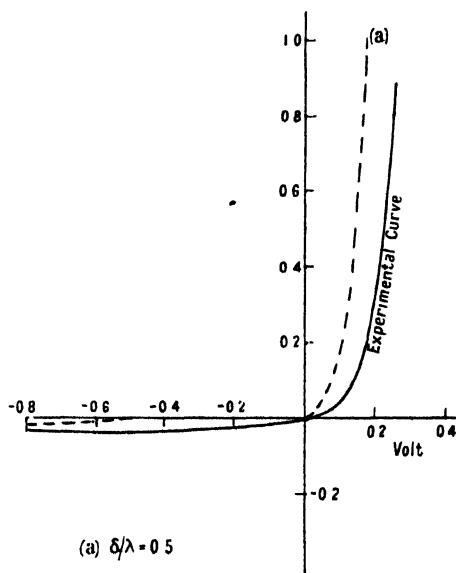


Figure 15. Curve showing rectification in the region V at the contact between germanium and a metal. Experimental data supplied by British Thomson-Houston Co. Ltd.

If the temperature T is changed, the thickness x_0 of the space-charge barrier, and therefore the critical ratio $\delta : x_0$ changes proportionally. Where the thermionic factor $e^{-\phi/kT}$ also enters, the total temperature effect is complex and is not considered here. For the germanium crystals, where δ may be taken as zero, one has more simply

$$i \propto \sigma e^{-\phi/kT} = \text{const.} \cdot e^{-E_0/2kT} e^{-\phi/kT},$$

as in Bethe's theory.

§ 4. CONCLUSION

The theory, applied in § 1 to the contact between identical semiconductors, accounts qualitatively for the behaviour of the contacts between a semiconductor and a metal, if the effect of space-charge is taken into account. It follows that both the Wilson and Schottky mechanisms are necessary to a complete treatment of crystal rectifiers.

REFERENCES

- BETHE, 1942, Theory of the boundary layer of crystal rectifiers, *B.L. Report*, No. 43-12.
 BRAUN and BUSCH, 1942, *Helv. Phys. Acta*, **15**, 571.
 FINCH and WILMAN, 1937, *Trans. Faraday Soc.*, **33**, 337.
 FOWLER and NORDHEIM, 1928, *Proc. Roy. Soc. A*, **119**, 173.
 HEINE and SCHERRER, 1940, *Helv. Phys. Acta*, **13**, 489.
 JEANS, *Electricity and Magnetism*, p. 358 (Cambridge University Press).
 KURTSCHATOW *et al.*, 1935, *Phys. Z. Sowjet*, **7**, 257.
 MITCHELL and SILLARS, 1947, *Metropolitan-Vickers Research Reports* (February).
 MOTT, 1939, *Proc. Roy. Soc. A*, **171**, 27.
 SCHOTTKY, 1939, *Wiss. Veröff. Siemens-Werke*, **18**, 225.
 WILSON, 1932, *Proc. Roy. Soc. A*, **136**, 487.

The Specific Heat of a Linear Ionic Lattice

By B. DONOVAN *

Imperial College, London

* Now at the University of Sheffield.

*Communicated by M. Blackman ; MS. received 12 June 1947 ;
 in amended form 16 September 1947*

ABSTRACT. The specific heat at moderately high temperatures has been calculated for a one-dimensional lattice in which both electrostatic and repulsive forces are taken into account. The Thirring expansion method, which does not necessitate a knowledge of the density of normal vibrations of the lattice, has been applied and the resulting values of the specific heat are compared with those obtained by graphical integration of the frequency spectrum.

Thirring's method involves expanding the specific heat in a series in which the terms include summations over the $2n$ th powers of the normal frequencies. It is found that values of n greater than 4 lead to computations which are excessively laborious. Nevertheless, with four terms, it is shown that reliable results may be obtained over a considerable range of temperature. At lower temperatures the Thirring method becomes progressively less accurate, showing that the number of terms used is insufficient, and a correction is derived from considerations of the remainder of the series after four terms.

§ 1. INTRODUCTION

THE properties of a one-dimensional lattice were first investigated by Born and von Kármán (1912) who worked out the distribution of normal vibrations on the assumption of quasi-elastic binding forces between neighbouring particles. The specific heat of such a lattice was evaluated by Blackman (1935), who compared his results with those derived from the continuum theory of Debye (1912). The case of the ionic lattice was examined by Broch (1936) who, using the generalized zeta-function, worked out the vibrational spectrum of a one-dimensional lattice taking into account the electrical interactions of the charged particles and the repulsive forces.

In this paper the specific heat of a linear ionic lattice is calculated by two methods. If the vibrational spectrum is represented by a function $\rho(\nu)$ giving the number of frequencies between ν and $\nu + d\nu$, the specific heat at constant volume can be obtained in the form

$$c_v = k \int_0^{\nu_{\max}} \rho(\nu) E\left(\frac{h\nu}{kT}\right) d\nu, \quad \dots\dots (1)$$

where

$$E(x) = \frac{x^2 e^x}{(e^x - 1)^2},$$

ν_{\max} is the maximum frequency, h is Planck's constant, k Boltzmann's constant and T the absolute temperature. Once the function $\rho(\nu)$ is known, the solution of (1) using graphical integration is relatively straightforward, and greater interest attaches to the application of a method due to Thirring (1913) by which the specific heat of moderately high temperatures may be obtained without a knowledge of the vibrational spectrum.

In this method the mean energy of a linear oscillator

$$\bar{u} = \frac{1}{2} h\nu + kT - \frac{h\nu/kT}{\exp(h\nu/kT) - 1},$$

is expanded in powers of $h\nu/kT$, giving

$$\bar{u} = kT \left\{ 1 - \sum_{n=1}^{\infty} (-1)^n \frac{B_n}{(2n)!} (h\nu/kT)^{2n} \right\},$$

where B_n are the Bernoulli numbers; the specific heat, obtained by differentiation, is thus expressible as a series of terms in ν^2 , ν^4 etc., which have to be summed over all the frequencies of the lattice. These terms become progressively more laborious to compute, and it is desirable to determine whether a comparatively small number of terms will yield useful results. In particular, it is important to know the temperature range over which the method is reliable, and the degree of accuracy to be expected. Some information may thus be obtained for the three-dimensional case.

We consider a chain of particles which may be taken to consist of a number of cells each containing two particles. The coordinate of the k th particle in the l th cell in the equilibrium configuration is given by

$$x_k^l = x_k + la,$$

where x_k is the coordinate of the k th particle in the base cell and a is the lattice constant. The distance between the k' th particle in the base cell and the k th particle in the l th cell is denoted by $x_{kk'}^l$, and the total potential energy of two

particles with coordinates x_k^l and $x_{k'}$ is denoted by $\phi_{kk'}^l$. The equations of motion then assume the form (Born 1923)

$$\left. \begin{aligned} \left\{ \omega^2 m_1 + \begin{bmatrix} 1 & 1 \\ x & x \end{bmatrix} \right\} U_1 + \begin{bmatrix} 1 & 2 \\ x & x \end{bmatrix} U_2 = 0, \\ \begin{bmatrix} 2 & 1 \\ x & x \end{bmatrix} U_1 + \left\{ \omega^2 m_2 + \begin{bmatrix} 2 & 2 \\ x & x \end{bmatrix} \right\} U_2 = 0. \end{aligned} \right\} \quad \dots\dots(2)$$

The symbols $\begin{bmatrix} k & k' \\ x & x \end{bmatrix}$ are defined by the equations

$$\begin{bmatrix} k & k' \\ x & x \end{bmatrix} = \sum_{l=-\infty}^{+\infty} (\phi_{kk'}^l)_{xx} \exp\{-2\pi i h x_{kk'}^l / a\}, \quad \dots\dots(3)$$

where $(\phi_{kk'}^l)_{xx} = \{d^2(\phi_{kk'}^l)/dx^2\} |x_{kk'}^l|$. \dots\dots(4)

Further, h/a is the wave-number, ω is the angular frequency, U_1 and U_2 are the amplitudes and m_1 and m_2 the masses of the two particles.

The potential $\phi_{kk'}^l$ may be conveniently resolved into two parts (Broch 1937), subsequently referred to as B.): one due to the Coulomb forces and the other due to the repulsive forces. The van der Waals attractive forces are neglected.

The functions $\begin{bmatrix} k & k' \\ x & x \end{bmatrix}$ are therefore similarly separable, and for the electrostatic parts one obtains (B., equations (10), (19) and (27)) :

$$\begin{bmatrix} 1 & 2 \\ x & x \end{bmatrix}^E = \begin{bmatrix} 2 & 1 \\ x & x \end{bmatrix}^E = -\frac{2\epsilon^2}{a^3} e^{\pi i h} \sum_{l=-\infty}^{+\infty} \frac{\exp(2\pi i h l)}{|l + \frac{1}{2}|^3} \quad \dots\dots(5)$$

and $\begin{bmatrix} 1 & 1 \\ x & x \end{bmatrix}^E = \begin{bmatrix} 2 & 2 \\ x & x \end{bmatrix}^E = \frac{24\epsilon^2}{a^3} \zeta(3) + \frac{2\epsilon^2}{a^3} \sum_{l=-\infty}^{+\infty} \frac{\exp(2\pi i h l)}{|l|^3},$
\dots\dots(6)

where the accent on the summation sign signifies the omission of the term $l=0$. Here $\zeta(t)$ is the Riemann zeta-function and the charge on the j th particle is $+\epsilon$ or $-\epsilon$ according as j is odd or even.

Assuming that the repulsive potential varies as the inverse s th power of the distance, and taking into account nearest neighbours only, the repulsive parts

$\begin{bmatrix} k & k' \\ x & x \end{bmatrix}^R$ may be expressed in the form (B., equations (34), (35), (37) and (48))

$$\begin{bmatrix} 1 & 2 \\ x & x \end{bmatrix}^R = \begin{bmatrix} 2 & 1 \\ x & x \end{bmatrix}^R = \frac{8\epsilon^2}{a^3} \cos \pi h (s+1) \log 2 \quad \dots\dots(7)$$

and $\begin{bmatrix} 1 & 1 \\ x & x \end{bmatrix}^R = \begin{bmatrix} 2 & 2 \\ x & x \end{bmatrix}^R = -\frac{8\epsilon^2}{a^3} (s+1) \log 2.$ \dots\dots(8)

Returning to the equations of motion (2), we see that the normal frequencies are obtained as roots of the determinantal equation

$$\begin{vmatrix} \omega^2 m_1 + A(h) & B(h) \\ B(h) & \omega^2 m_2 + A(h) \end{vmatrix} = 0, \quad \dots\dots(9)$$

where $A(h) = \frac{8\epsilon^2}{a^3} [3\zeta(3) - (s+1) \log 2] + \frac{2\epsilon^2}{a^3} \sum_{l=-\infty}^{+\infty} \frac{\exp(2\pi i h l)}{|l|^3}$ \dots\dots(10)

and $B(h) = \frac{8\epsilon^2}{a^3} \cos \pi h (s+1) \log 2 - \frac{2\epsilon^2}{a^3} e^{\pi i h} \sum_{l=-\infty}^{+\infty} \frac{\exp(2\pi i h l)}{|l + \frac{1}{2}|^3}.$ \dots\dots(11)

In order to consider the distribution of the frequencies we specialize to a chain consisting of N cells, i.e. of $2N$ particles. Application of periodic boundary conditions then gives

$$\exp(2\pi i N h) = 1.$$

Now $0 \leq h < 1$ (since the wavelength cannot be less than the lattice constant); consequently we have

$$h = \mu/N, \quad \dots\dots(12)$$

where μ is any positive integer such that $0 < \mu \leq N$. The allowed values of h have a constant spacing ($= 1/N$) and the normal frequencies are distributed uniformly in h -space, each value of h corresponding to two frequencies.

In the method of Thirring (1913) the heat capacity of a linear oscillator (of frequency ν) is expressed in the form

$$c_\nu = k \left\{ 1 + \sum_{n=1}^{\infty} (-1)^n \frac{B_n(2n-1)}{(2n)!} \left(\frac{h\nu}{kT} \right)^{2n} \right\},$$

and the series is convergent if $h\nu/kT < 2\pi$. In this expression B_n are the Bernoulli numbers.

In order to derive the heat capacity of the one-dimensional lattice, it is necessary to sum this expression over all the $2N$ frequencies of the lattice, i.e.

$$c_\nu = k \left\{ 2N + \sum_{n=1}^{\infty} (-1)^n \frac{B_n 2n-1}{(2n)!} \left(\frac{h}{kT} \right)^{2n} \sum_{\nu} \nu^{2n} \right\}. \quad \dots\dots(13)$$

We thus have to perform the summations over all the normal frequencies using equations (9), (10) and (11). For the case in which the two types of particle have equal masses, we write $m_1 = m_2 = m_0$ and the solutions of (9) become

$$\left. \begin{aligned} 4\pi^2 m_0 \nu_1^2 &= -A(h) - B(h), \\ 4\pi^2 m_0 \nu_2^2 &= -A(h) + B(h). \end{aligned} \right\} \quad \dots\dots(14)$$

From these solutions the sums $\sum_{\nu} \nu^{2n}$ may be evaluated, and the cases corresponding to $n = 1, 2, 3, 4$ are treated individually in the next four sections. For $n > 4$ the labour involved is prohibitive and a discussion of the error incurred in neglecting higher order terms is given later. The equations (14) show how two frequencies are associated with every value of h ; the frequency curve splits into two branches but these two solutions are identical apart from a difference in phase (B., p. 493).

§ 2. EVALUATION OF $\sum_{\nu} \nu^2$

From (14) we obtain

$$\nu_1^2 + \nu_2^2 = -(1/2\pi^2 m_0) A(h),$$

and, summing over both branches to include all frequencies, we can express the required sum as

$$\sum_{\nu} \nu^2 = -(1/2\pi^2 m_0) \sum_h A(h), \quad \dots\dots(15)$$

where the right-hand side is to be summed over all the permissible values of h . Now from (12) it is clear that the summation over h is equivalent to a summation over μ and, since N is very large, this may be replaced by the integral

$$\int_0^N A(\mu/N) d\mu, \quad \text{i.e. by} \quad \int_0^1 N A(h) dh.$$

Reference to (10) shows that $A(h)$ consists of two parts, the first independent of h and the second an infinite series involving h . The latter is uniformly convergent in the interval $0 \leq h \leq 1$ and may therefore be integrated term by term. This, however, gives zero for all integral values of l and we are thus left with

$$\Sigma \nu^2 = -(4N\epsilon^2/\pi^2 m_0 a^3) [3\zeta(3) - (s+1) \log 2]. \quad \dots\dots (16)$$

To obtain the numerical value of the expression in square brackets (subsequently denoted by C) the exponent s of the repulsive forces is given the value 10. Further,

$$\zeta(3) = \sum_{n=1}^{\infty} 1/n^3 = 1.20205 \dots,$$

$$\text{so that} \quad C = -4.01845. \quad \dots\dots (17)$$

Substituting in (16) gives

$$\Sigma \nu^2 = 1.629 N(\epsilon^2/m_0 a^3). \quad \dots\dots (18)$$

§ 3. EVALUATION OF $\Sigma \nu^4$

From (14) we obtain

$$\Sigma \nu^4 = N/8\pi^4 m_0^2 \left[\int_0^1 A^2(h) dh + \int_0^1 B^2(h) dh \right]. \quad \dots\dots (19)$$

$$\text{Now} \quad A^2(h) = 64\epsilon^4 a^{-6} C^2 + 32\epsilon^4 a^{-6} C \sigma_1 + 4\epsilon^4 a^{-6} \sigma_1^2, \quad \dots\dots (20)$$

$$\text{where} \quad \sigma_1 = \sum_l \frac{\exp(2\pi i h l)}{|l|^3}.$$

Of these three terms the first is independent of h , the second provides a zero contribution and the third may be expressed as a double sum by means of Cauchy's theorem concerning the multiplication of absolutely convergent series. Upon integration the non-zero terms reduce to

$$\sum_{n=-\infty}^{+\infty} \sum_{n'=1}^{\sigma} 1/n^6 = 2 \sum_{n=1}^{\sigma} 1/n^6,$$

$$\text{and, since} \quad \sum_{n=1}^{\infty} 1/n^6 = 1.01734 \dots,$$

$$\text{we have} \quad \int_0^1 A^2(h) dh = 1.042 \times 10^3 \epsilon^4/a^6. \quad \dots\dots (21)$$

Squaring expression (11) gives

$$B^2(h) = 64\epsilon^4 a^{-6} [(s+1) \log 2]^2 \cos^2 \pi h - 32\epsilon^4 a^{-6} (s+1) \log 2 \cos \pi h \cdot \sigma_2 + 4\epsilon^4 a^{-6} \cdot \sigma_2^2, \quad \dots\dots (22)$$

$$\text{where} \quad \sigma_2 = \sum_l \frac{\exp[\pi i h (2l+1)]}{|l+\frac{1}{2}|^3}.$$

Here the first term is straightforward, and the second splits into two sums both of which give, when integrated, a numerical value of 8. Application of Cauchy's theorem to the third term in (22) leads to the expression

$$\sum_{n=-\infty}^{+\infty} \frac{1}{|n+\frac{1}{2}|^6} = 2 \sum_{n=0}^{\infty} \frac{1}{(n+\frac{1}{2})^6} = 126 \zeta(6),$$

using (71) with $\alpha=6$.

Hence
$$\int_0^1 B^2(h) dh = 421.2 \epsilon^4 / a^6, \quad \dots\dots(23)$$

and, inserting (21) and (23) into (19), we have

$$\Sigma \nu^4 = 1.877 N (\epsilon^2 / m_0 a^3)^2. \quad \dots\dots(24)$$

§ 4. EVALUATION OF $\Sigma \nu^6$

From (14) we obtain

$$\Sigma \nu^6 = - (N / 32 \pi^6 m_0^3) \left[\int_0^1 A^3(h) dh + 3 \int_0^1 A(h) B^2(h) dh \right]. \quad \dots\dots(25)$$

Now

$$A^3(h) = 512 \epsilon^6 a^{-9} C^3 + 384 \epsilon^6 a^{-9} C^2 \sigma_1 + 96 \epsilon^6 a^{-9} C \sigma_1^2 + 8 \epsilon^6 a^{-9} \sigma_1^3. \quad \dots\dots(26)$$

Here the first term is constant, the second vanishes when integrated, and the value of the third may be obtained from § 3. The fourth term reduces to the double sum

$$\sum_m' \sum_n' [|m+n|^3 |m|^3 |n|^3]^{-1}, \quad (m+n \neq 0),$$

which is shown in Appendix 1 to have the numerical value 0.816. Integration of (26) thus gives

$$\int_0^1 A^3(h) dh = -3.400 \times 10^4 \epsilon^6 / a^9. \quad \dots\dots(27)$$

For the second term in (25) we find

$$\begin{aligned} A(h) B^2(h) = & 512 \epsilon^6 a^{-9} C [(s+1) \log 2]^2 \cos^2 \pi h + 32 \epsilon^6 a^{-9} C \sigma_2^2 \\ & - 256 \epsilon^6 a^{-9} C (s+1) \log 2 \cos \pi h \cdot \sigma_2 + 128 \epsilon^6 a^{-9} [(s+1) \log 2]^2 \cos^2 \pi h \cdot \sigma_1 \\ & + 8 \epsilon^6 a^{-9} \sigma_1 \sigma_2^2 - 64 \epsilon^6 a^{-9} (s+1) \log 2 \cos \pi h \cdot \sigma_1 \sigma_2. \end{aligned} \quad \dots\dots(28)$$

Of these six terms the first is straightforward and the second and third may be evaluated using the results of § 3. The fourth term splits into three sums: one gives a zero contribution and the other two have the same value, namely unity. The fifth term may be expanded as a triple sum which reduces to

$$\sum_m \sum_n [|m+n+1|^3 |m+\frac{1}{2}|^3 |n+\frac{1}{2}|^3]^{-1}, \quad (m+n \neq -1),$$

and by adding the fifty largest terms a numerical value of 139 is obtained.

The sixth term in (28) may be written as two double sums which lead to the expression

$$4 \left[\sum_m' [|m+1|^3 |2m+1|^3]^{-1} + \sum_m' [|m|^3 |2m+1|^3]^{-1} \right].$$

It is easily seen that each of these terms is equivalent to

$$\sum_{m=1}^{\infty} [m^3 (2m+1)^3]^{-1} + \sum_{m=1}^{\infty} [m^3 (2m-1)^3]^{-1},$$

and the value of this combination (obtained from the general case in Appendix 2) is

$$80 - 48 \zeta(2),$$

i.e. $1.0432 \quad (\zeta(2) = 1.6449 \dots).$

Integration of (28) thus gives

$$\int_0^1 A(h) B^2(h) dh = -1.278 \times 10^4 \epsilon^6 / a^9, \quad \dots\dots(29)$$

and substituting (27) and (29) into (25) yields

$$\Sigma \nu^6 = 2.351 N (\epsilon^2 / m_0 a^3)^3. \quad \dots\dots(30)$$

§ 5. EVALUATION OF $\Sigma \nu^4$

From (14) we obtain

$$\Sigma \nu^4 = (N/128\pi^8 m_0^4) \left[\int_0^1 A^4(h) dh + 6 \int_0^1 A^2(h) B^2(h) dh + \int_0^1 B^4(h) dh \right]. \quad \dots\dots(31)$$

Now

$$A^4(h) = 4096\epsilon^8 a^{-12} C^4 + 4096\epsilon^8 a^{-12} C^3 \sigma_1 + 1536\epsilon^8 a^{-12} C^2 \sigma_1^2 + 256\epsilon^8 a^{-12} C \sigma_1^3 + 16\epsilon^8 a^{-12} \sigma_1^4. \quad \dots\dots(32)$$

Here the first term is constant, the second vanishes when integrated, the third is evaluated in § 3 and the fourth in § 4 (using Appendix 1). The fifth term reduces to the triple sum

$$\sum_{l,m,n}' \sum_{l,m,n}' [|l|^3 |m|^3 |n|^3 |l+m+n|^3]^{-1}, \quad (l+m+n \neq 0),$$

and addition of the forty largest terms gives a numerical value of 6.79. Integration of (32) thus gives

$$\int_0^1 A^4(h) dh = 11.178 \times 10^5 \epsilon^8 / a^{12}. \quad \dots\dots(33)$$

The second term in (31) is obtained from

$$\begin{aligned} A^2(h) B^2(h) = & 4096\epsilon^8 a^{-12} C^2 [(s+1) \log 2]^2 \cos^2 \pi h \\ & + 256\epsilon^8 a^{-12} C^2 \sigma_2^2 - 2048\epsilon^8 a^{-12} C^2 (s+1) \log 2 \cos \pi h \cdot \sigma_2 \\ & + 256\epsilon^8 a^{-12} [(s+1) \log 2]^2 \cos^2 \pi h \cdot \sigma_1^2 + 16\epsilon^8 a^{-12} \sigma_1^2 \sigma_2^2 \\ & - 128\epsilon^8 a^{-12} (s+1) \log 2 \cos \pi h \cdot \sigma_2 \sigma_1^2 \\ & + 2048\epsilon^8 a^{-12} C [(s+1) \log 2]^2 \cos^2 \pi h \cdot \sigma_1 + 128\epsilon^8 a^{-12} C \sigma_1 \sigma_2^2 \\ & - 1024\epsilon^8 a^{-12} C (s+1) \log 2 \cos \pi h \cdot \sigma_1 \sigma_2. \end{aligned} \quad \dots\dots(34)$$

Of these nine terms the first may be integrated straightforwardly and the second and third evaluated using the results of § 3. The fourth term splits into three double sums which lead to the expression

$$\frac{1}{4} \left[\sum_m' [|m|^3 |m+1|^3]^{-1} + \sum_m' [|m|^3 |m-1|^3]^{-1} + 2 \sum_m' |m|^{-6} \right].$$

This reduces to $\sum_m' [m^3(m+1)^3]^{-1} + \sum_m' m^{-6}$ (cf. equation (63))

and, using the results of Appendix 1, the numerical values are found to be

$$0.1304 + 1.0173 = 1.1477.$$

The fifth term in (34) may be expanded as a quadruple sum which reduces to

$$\sum_1' \sum_m' \sum_n' [|l|^3 |m|^3 |n + \frac{1}{2}|^3 |l+m+n + \frac{1}{2}|^3]^{-1},$$

and by adding the forty largest terms a numerical value of 309 is obtained. The sixth term in (34) splits into two triple sums and gives

$$4 \left[\sum_m' \sum_n' [|m|^3 |n|^3 |2m+2n+1|^3]^{-1} + \sum_m' \sum_n' [|m|^3 |n|^3 |2m+2n-1|^3]^{-1} \right],$$

in which both terms have the same value, since changing the signs of m and n in either sum gives the other. By direct addition of the forty largest terms this value is found to be 2.345.

The seventh, eighth and ninth terms in (34) are similar to the fourth, fifth and sixth terms respectively in (28) and hence integration of (34) gives

$$\int_0^1 A^2(h) B^2(h) dh = 3.900 \times 10^5 \epsilon^8 / a^{12}. \quad \dots\dots (35)$$

Lastly we consider the third term in (31), which is obtained from

$$\begin{aligned} B^4(h) = & 4096\epsilon^8 a^{-12} [(s+1) \log 2]^4 \cos^4 \pi h - 4096\epsilon^8 a^{-12} [(s+1) \log 2]^3 \cos^3 \pi h \cdot \sigma_2 \\ & + 1536\epsilon^8 a^{-12} [(s+1) \log 2]^2 \cos^2 \pi h \cdot \sigma_2^2 - 256\epsilon^8 a^{-12} (s+1) \log 2 \cos \pi h \cdot \sigma_2^3 \\ & + 16\epsilon^8 a^{-12} \sigma_2^4. \end{aligned} \quad \dots\dots (36)$$

The first term here is straightforward and the second, written as four separate sums, may be shown to have the value

$$1/8 [8/27 + 24 + 24 + 8/27] = 6.074.$$

The third term may be expressed as three double sums which reduce in the usual way to

$$16 \left[\sum_l [|2l+1|^3 |2l+3|^3]^{-1} + \sum_l [|2l+1|^3 |2l-1|^3]^{-1} + 2 \sum_l |2l+1|^{-6} \right]. \quad \dots\dots (37)$$

In this expression the first two terms have the same value (the second emerges from the first on replacing l by $-(l+1)$).

Now
$$\sum_l [|2l+1|^3 |2l+3|^3]^{-1} = 1 + 2 \sum_{l=1}^{\infty} (2l-1)^{-3} (2l+1)^{-3},$$

and
$$\sum_{l=1}^{\infty} (2l-1)^{-3} (2l+1)^{-3} = \sum_{l=1}^{\infty} (4l^2-1)^{-3} = (32-3\pi^2)/64 = 0.03736$$

(Knopp 1928, p. 269).

Further, from (71)

$$\sum_l |2l+1|^{-6} = 63 \zeta(6)/32,$$

so that the numerical value of (37) is 98.483.

The fourth term in (36) splits into two triple sums which lead to

$$\begin{aligned} 256 \left[\sum_m \sum_n [|2m+1|^3 |2n+1|^3 |2m+2n+3|^3]^{-1} \right. \\ \left. + \sum_m \sum_n [|2m+1|^3 |2n+1|^3 |2m+2n+1|^3]^{-1} \right], \end{aligned}$$

and by adding the forty largest terms the value of each sum is found to be 3.12.

To evaluate the fifth term in (36) we have to consider the expression

$$2^{12} \sum_l \sum_m \sum_n [|2l+1|^3 |2m+1|^3 |2n+1|^3 |2l+2m+2n+3|^3]^{-1},$$

and by direct addition, ignoring all terms in which l, m, n are greater than unity, an approximate value of 2.86×10^4 is obtained.

Integration of (36) thus gives

$$\int_0^1 B^4(h) dh = 2.957 \times 10^5 \epsilon^8 / a^{12}, \quad \dots\dots (38)$$

and substituting (33), (35) and (38) into (31) finally gives

$$\Sigma \nu^8 = 3.091 N(\epsilon^2/m_0 a^3)^4. \quad \dots\dots (39)$$

§ 6. CALCULATION OF SPECIFIC HEAT BY THIRRING'S METHOD

The specific heat, which may now be computed from (13) by inserting the results given by (18), (24), (30) and (39), is given by

$$c_v = Nk \left\{ 2 - \left(\frac{1.629}{12} \right) \Lambda + \left(\frac{1.877}{240} \right) \Lambda^2 - \left(\frac{2.351}{6048} \right) \Lambda^3 + \left(\frac{3.091}{172800} \right) \Lambda^4 \right\}, \quad \dots (40)$$

where $\Lambda = (\hbar/kT)^2 (\epsilon^2/m_0 a^3)$.

Since values of the specific heat obtained from this formula are to be compared with the results of graphical integration it is essential to ensure that the maximum frequency is the same in both cases. The angular frequency may be expressed in terms of the wave number by means of the relation (B, p. 493)

$$\begin{aligned} m_0 \omega^2 = & 8\epsilon^2 a^{-3} (s+1) \log 2(1 - \cos \pi h) - 24\epsilon^2 a^{-3} \zeta(3) \\ & - 4\epsilon^2 a^{-3} \left[\frac{2\pi}{3} + 2 \sum_{l=1}^{\infty} \left[\frac{\exp(-\pi l^2)}{l^2} + \frac{G(\sqrt{\pi}l)}{2l^3} \right] \cos 2\pi h l \right. \\ & \left. + \pi \sum_{l=-\infty}^{+\infty} [\exp(-\pi[h+l]^2) + \pi[h+l]^2 \text{Ei}(-\pi[h+l]^2)] \right\} \\ & + 4\epsilon^2 a^{-3} \left\{ 2 \sum_{l=0}^{\infty} \left[\frac{\exp(-\pi[\frac{1}{2}+l]^2)}{(\frac{1}{2}+l)^2} + \frac{G(\sqrt{\pi}[\frac{1}{2}+l])}{2(\frac{1}{2}+l)^3} \right] \cos 2\pi h(\frac{1}{2}+l) \right. \\ & \left. + \pi \sum_{l=-\infty}^{+\infty} (-1)^l [\exp(-\pi[h+l]^2) + \pi[h+l]^2 \text{Ei}(-\pi[h+l]^2)] \right\} \end{aligned} \quad (41)$$

In this expression $G(x) = 1 - F(x)$,

where $F(x)$ is the Gauss error integral, and $\text{Ei}(-x)$ is defined by

$$-\text{Ei}(-x) = \int_x^{\infty} \frac{e^{-t}}{t} dt \quad (0 < x < \infty).$$

The ω - h curve (B., p. 498) shows that there is a maximum value for the frequency when h is approximately 0.88*; the density of normal vibrations (given by $dh/d\omega$) therefore becomes infinite at this point. The behaviour of the density function in the neighbourhood of the maximum frequency, though not troublesome from the physical point of view, provides difficulties in graphical integration, since it is not possible in this region to estimate accurately the area under the density curve.

To avoid this difficulty we approximate to the exact frequency spectrum by replacing the high-frequency end by a single "line" (i.e. by an Einstein term) whose frequency is equal to that corresponding to $h = 1$. Figure 1 shows this spectrum together with the exact relation according to equation (41). The range of h over which the Einstein term extends is decided by reference to the density curve, and the upper limit for the graphical integration has been taken to correspond to $h = 0.675$.

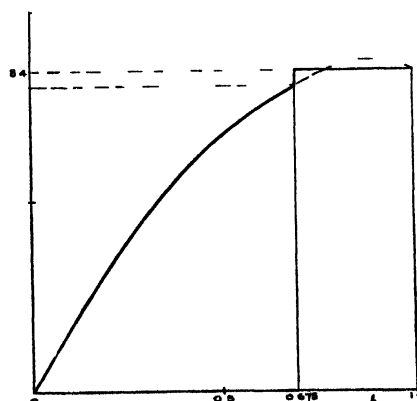


Figure 1. The angular frequency ω as a function of h . The scales are identical with those of B, figure 1.

* The exact position of the maximum depends on the exponent s of the repulsive forces.

The frequency ν_m of the Einstein term and the maximum frequency ν_0 of the continuous spectrum may be determined from (41). The series in this expression are rapidly convergent and numerical calculation is a fairly simple matter. Further, the functions $F(x)$ and $Ei(-x)$ are both available in numerical tables.* The calculations give

$$\nu_m = 1.1769(\epsilon^2/(m_0 a^3))^{\frac{1}{2}}, \quad \dots\dots(42)$$

$$\nu_0 = 1.1208(\epsilon^2/(m_0 a^3))^{\frac{1}{2}}. \quad \dots\dots(43)$$

In the Thirring calculation the maximum frequency is adjusted to have the value given by (43). At this stage the parameter θ_0 is introduced, defined by

$$\theta_0 = h\nu_0/k, \quad \dots\dots(44)$$

and the variation of the specific heat with temperature is obtained by plotting c_v as a function of θ_0/T . From (43) and (44) we have

$$\left(\frac{h}{kT}\right)^2 \left(\frac{\epsilon^2}{m_0 a^3}\right) = \left(\frac{\theta_0}{1.121T}\right)^2, \quad \dots\dots(45)$$

and this expression, substituted in (40), enables the specific heat to be calculated for any value of θ_0/T .

Table 1 shows the values obtained by this method over a range of temperature extending down to $\theta_0/T = 4.0$. The table also shows the results of the graphical method discussed in § 8. The results are shown graphically in figures 2 and 3, in which the specific heat is plotted against θ_0/T and T/θ_0 respectively. At high temperatures the specific heat tends to the classical value $2Nk$. For values of θ_0/T up to 3.0 excellent agreement is obtained between the two methods, the maximum discrepancy (approximately 0.5%) occurring in the neighbourhood of $\theta_0/T = 0.7$. For $\theta_0/T > 3.0$ the Thirring curve diverges from the graphical curve and gives higher values for the specific heat. This indicates that the formula (40)



Figure 2. The specific heat as a function for θ_0/T . The corrections given in § 7 are shown for $\theta_0/T = 3.0, 3.5$ and 4.0 .

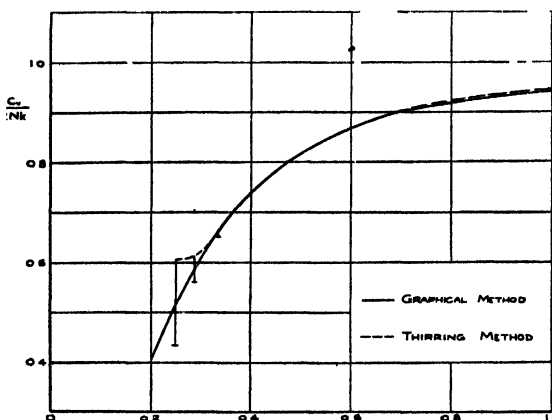


Figure 3. The specific heat as a function of T/θ_0 . The corrections given in § 7 are shown for $\theta_0/T = 3.0, 3.5$ and 4.0 .

* e.g. Publications of the Federal Works Agency, Works Projects Administration (New York 1940-41).

is valid up to $\theta_0/T = 3.0$; above this value the results become rapidly less accurate, as might be expected, since the infinite series in (13) has been cut off at the fourth term.

Table 1

θ_0/T	0	0.3	0.7	1.0	1.25	1.5	2.0	2.5	3.0	3.5	4.0	5.0
$c_v/2Nk$ (Thirring)	1.000	0.995	0.974	0.948	0.921	0.890	0.818	0.741	0.667	0.611	0.604	
$c_v/2Nk$ (graphical)	1.000	0.993	0.969	0.945	0.918	0.892	0.817	0.740	0.667	0.587	0.519	0.407

§ 7. CONSIDERATION OF REMAINDER IN THIRRING'S METHOD

Having found the range of θ_0/T over which equation (40) can yield useful results, it is of interest to investigate the remainder after four terms in the original expression (13). In this section a rough estimate is made of the upper limit to the remainder in terms of θ_0/T .

An extension of Euler's summation formula is used, starting with the identity (Knopp 1928, p. 528)

$$e^{\alpha} - 1 + \frac{\alpha}{2} - 1 - \sum_{n=1}^{\lambda} (-1)^n \frac{B_n}{(2n)!} \alpha^{2n} + \frac{\alpha^{2\lambda+2}}{e^{\alpha} - 1} \int_0^1 P_{2\lambda+1}(x) e^{\alpha x} dx, \quad \dots\dots (46)$$

$$\text{where} \quad P_{2\lambda+1}(x) = (-1)^{\lambda-1} \sum_{m=1}^{\lambda} (2 \sin 2m\pi x) / (2m\pi)^{2\lambda+1}. \quad \dots\dots (47)$$

Now the mean energy of a linear oscillator is given by

$$\bar{u} = \frac{1}{2} h\nu + h\nu / [\exp(h\nu/kT) - 1],$$

and hence the identity (46), with $\alpha = h\nu/kT$, shows that the remainder after λ terms is

$$kT \left\{ \alpha^{2\lambda+2} (e^{\alpha} - 1)^{-1} \int_0^1 P_{2\lambda+1}(x) e^{\alpha x} dx \right\}.$$

Differentiating this expression we obtain the remainder $R_{\lambda} =$ in the expansion for the specific heat:

$$\begin{aligned} R_{\lambda} = & k\alpha^{2\lambda+2} (e^{\alpha} - 1)^{-1} \int_0^1 P_{2\lambda+1}(x) e^{\alpha x} dx \\ & - (h\nu/T) [(2\lambda+2)(e^{\alpha} - 1)\alpha^{2\lambda+1} - e^{\alpha}\alpha^{2\lambda+2}] (e^{\alpha} - 1)^{-2} \int_0^1 P_{2\lambda+1}(x) e^{\alpha x} dx \\ & - (h\nu/T) \alpha^{2\lambda+2} (e^{\alpha} - 1)^{-1} \int_0^1 x P_{2\lambda+1}(x) e^{\alpha x} dx. \quad \dots\dots (48) \end{aligned}$$

To determine the maximum value of $|R_{\lambda}|$, with $\lambda = 4$, we consider the first two terms together and write them as

$$k \{ \alpha^{10} (e^{\alpha} - 1)^{-1} - (10\alpha^{10} (e^{\alpha} - 1) - \alpha^{11} e^{\alpha}) (e^{\alpha} - 1)^{-2} \} \int_0^1 P_9(x) e^{\alpha x} dx. \quad \dots\dots (49)$$

An upper limit to the value of the integral may be derived from the inequality

$$|P_9(x)| \leq (2/(2\pi)^9) \sum_{m=1}^{\infty} m^{-9},$$

which follows from (47). Since

$$\sum_{m=1}^{\infty} m^{-9} < \sum_{m=1}^{\infty} m^{-8}, \quad \text{and} \quad \sum_{m=1}^{\infty} m^{-8} = \pi^8/9450,$$

$|P_9(x)|$ is certainly less than the quantity \hat{P} , where

$$\hat{P} = (256 + 9450\pi)^{-1}. \quad \dots\dots (50)$$

Thus for the limiting value of $\int_0^1 P_9(x)e^{\alpha x} dx$ we may write

$$\hat{P} \int_0^1 e^{\alpha x} dx = \hat{P} \alpha^{-1} (e^{\alpha} - 1),$$

which, substituted in (49), gives

$$k\hat{P}\{-9\alpha^9 + \alpha^{10} + \alpha^{10}/(e^{\alpha} - 1)\}. \quad \dots\dots (51)$$

Since the last term contains $e^{\alpha} - 1$ in the denominator, we may write $e^{\alpha} - 1 = \alpha$ to obtain the form

$$k\hat{P}\alpha^9(\alpha - 8),$$

which indicates that (51) is negative, since the convergence of formula (13) requires $\alpha < 6$.

We now examine the third term in (48), viz.

$$k\alpha^{11}(e^{\alpha} - 1)^{-1} \int_0^1 xP_9(x)e^{\alpha x} dx, \quad \dots\dots (52)$$

and seek to establish the sign of the integral since this term has to be subtracted from the first two. The function $P_9(x)$ has the form sketched in figure 4 (Knopp 1928, p. 537).

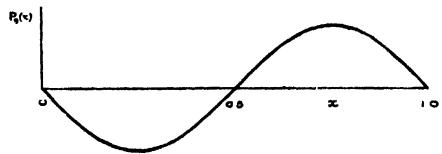


Figure 4. Form of the function $P_9(x)$.

In view of the fact that the integral contains the factor $e^{\alpha x}$, where α is positive, it is clear that the integral itself, and hence the expression (52), must be positive.

To obtain the maximum value of $|R|_{\lambda=4}$, therefore, the maximum value of (52) must be subtracted from (51). The limiting value of the integral in (52) may be expressed as

$$\hat{P} \int_0^1 x e^{\alpha x} dx = \hat{P} \alpha^{-2} \{e^{\alpha}(\alpha - 1) + 1\},$$

which, substituted in (52), gives

$$k\hat{P}\{\alpha^{10}e^{\alpha}(e^{\alpha} - 1)^{-1} - \alpha^9\}.$$

Subtracting this expression from (51) we obtain $-8k\hat{P}\alpha^9$ so that we may write

$$|R| < 8k\hat{P}(h\nu/kT)^9,$$

which, summed over all possible frequencies, becomes

$$|R| < 8k\hat{P}(h/kT)^9 \Sigma \nu^9. \quad \dots\dots (53)$$

Now $\Sigma \nu^9 < \nu_0 \Sigma \nu^8$ and hence, using (39) and (43), we can define a quantity $|R|_{\max.}$ by the equation

$$|R|_{\max.} = 8N(1.12 \times 3.09)k\hat{P}(\hbar/kT)^9(\epsilon^2/m_0a^3)^{9/2}.$$

Finally we substitute from (45) and (50) and obtain

$$(|R|_{\max.}/2Nk) = (1.82 \times 10^{-6})[\theta_0/1.12T]^9. \quad \dots\dots(54)$$

Since R has been shown to be negative, this equation gives an upper limit to the correction which must be subtracted from the values given in table 1.

Values of the remainder calculated from (54) are given in table 2 and the corrections are indicated by vertical lines in figures 2 and 3.

The correction increases rapidly with decreasing temperature, since it is proportional to the ninth power of θ_0/T , and although negligible for $\theta_0/T < 2.0$ it is nearly 2% at $\theta_0/T = 3.0$. This is not borne out by the results obtained (table 1) and it is clear from the somewhat drastic approximations that the upper limit to $|R|$ given by (54) is considerably in excess of the true correction. For the two cases $\theta_0/T = 3.5$ and 4.0 the values given in table 2 are approximately twice the discrepancies between the "Thirring" and "Graphical" curves in figures 2 and 3.

Table 2

θ_0/T	1.0	2.0	3.0	3.5	4.0
$ R _{\max.}/2Nk$	5×10^{-7}	3×10^{-4}	0.01 ₃	0.05 ₁	0.17 ₁

§ 8. CALCULATION OF SPECIFIC HEAT BY GRAPHICAL INTEGRATION

The fundamental expression (1) may be evaluated by graphical means if the density of normal vibrations is known. In the present case, with the frequency spectrum illustrated in figure 1, the specific heat may be expressed in the form

$$c_v = k \int_0^{\nu_0} \rho(\nu) E\left(\frac{\hbar\nu}{kT}\right) d\nu + 2Nk t_E E\left(\frac{\hbar\nu_m}{kT}\right), \quad \dots\dots(55)$$

where the function $\rho(\nu)$ is normalized by means of the relation

$$\int_0^{\nu_0} \rho(\nu) d\nu = 2Nt_c. \quad \dots\dots(56)$$

Here t_E and t_c are weight factors which represent the fractions of the total number of vibrations included in the Einstein and continuous regions respectively. Since equal ranges of h contain equal numbers of vibrations, we have

$$t_c = 0.675, \quad t_E = 0.325. \quad \dots\dots(57)$$

The frequency spectrum for the region $0 \leq \nu \leq \nu_0$ has been calculated from equation (41) and the values of ω and h so obtained have been used to calculate the density function $\rho(\nu)$ in terms of ω by means of the relation (B, p. 499)

$$\begin{aligned} \frac{\omega}{\rho} = & \frac{4\pi\epsilon^2}{a^3m_0} \left\{ (s+1) \log 2 \sin \pi h + 2 \sum_{l=1}^{\infty} \left[\frac{\exp(-\pi l^2)}{l} + \frac{G(\sqrt{\pi}l)}{2l^2} \right] \sin 2\pi h l \right. \\ & - \pi \sum_{l=-\infty}^{+\infty} [1 - (-1)^l](h+l) \text{Ei}(-\pi[h+l]^2) \\ & \left. - 2 \sum_{l=0}^{\infty} \left[\frac{\exp(-\pi[\frac{1}{2}+l]^2)}{\frac{1}{2}+l} + \frac{G(\sqrt{\pi}[\frac{1}{2}+l])}{2(\frac{1}{2}+l)^2} \right] \sin 2\pi h(\frac{1}{2}+l) \right\}. \quad \dots\dots(58) \end{aligned}$$

For any value of θ_0/T we find can $E(h\nu/kT)$ as a function of ω ; computation is facilitated by means of the relation

$$E(x) = x^2 e^x / (e^x - 1)^2 = \frac{1}{2} x^2 / (\cosh x - 1). \quad \dots\dots (59)$$

Hence the product $\rho(\nu)E(h\nu/kT)$ can be obtained as a function of ω . In this way curves have been constructed for values of θ_0/T varying from 0.3 to 5.0, and the areas under the curves determined by counting squares. The second term in (55) may be calculated immediately from (59), since

$$\theta_m/T = 1.05 \theta_0/T,$$

and the sum of the two contributions gives the specific heat, as listed in table 1.

Referring to figures 2 and 3, the slight discrepancy in the neighbourhood of $\theta_0/T = 0.7$ is rather puzzling, but could probably be decreased by raising the frequency ν_m of the Einstein term or increasing its weight t_E . At the low temperature end of the curves one would expect the graphical method to give reasonably accurate results, since in this region the specific heat is controlled by the low-frequency end of the spectrum and hence the effect of the approximation (figure 1) is least noticeable.

ACKNOWLEDGMENT

The author wishes to thank Dr. M. Blackman for suggesting this problem and for his continued interest in it, and for many helpful discussions during the course of the work.

APPENDIX 1

In this appendix a method is given by which an approximate numerical value may be obtained for the expression

$$S_1 = \sum_m' \sum_n' [|m+n|^3 |m|^3 |n|^3]^{-1} \quad \dots\dots (60)$$

$$(m \neq 0, \quad n \neq 0, \quad m \neq -n).$$

Considering first positive values of n only, and putting $n = \alpha$, we obtain series of the form

$$\alpha^{-3} \sum_m' [|m|^3 |m+\alpha|^3]^{-1},$$

and in the two regions $m > 0$ and $m < -\alpha$, we have

$$\alpha^{-3} \sum_{m=-1}^{\infty} m^{-3} (m+\alpha)^{-3}, \quad \dots\dots (61)$$

while between the two prohibited values of m (0 and $-\alpha$) there are $(\alpha-1)$ terms, whose sum may be represented by

$$\sum \xi^{-3} \eta^{-3}, \quad \dots\dots (62)$$

taken over all non-zero values of ξ and η satisfying the condition $\xi + \eta = \alpha$.

The same reasoning will apply to negative values of n . For if we write $n = -\alpha$ and consider the series

$$\alpha^{-3} \sum_m' [|m|^3 |m-\alpha|^3]^{-1},$$

it is clear that the two regions $m > \alpha$ and $m < 0$ will each yield (61) and the remaining $(\alpha-1)$ terms will be given, as before, by (62).

We thus have the result

$$\sum_m' [|m|^3 |m + \alpha|^3]^{-1} + \sum_m' [|m|^3 |m - \alpha|^3]^{-1} = 4 \sum_{m=1}^{\infty} m^{-3} (m + \alpha)^{-3} + 2 \sum \xi^{-3} \eta^{-3}, \quad \dots\dots (63)$$

which we can apply to (60) and obtain

$$S_1 = 4 \sum_{m=1}^{\infty} [m^{-3} (m + 1)^{-3}] + \frac{1}{8} \left[2 + 4 \sum_{m=1}^{\infty} m^{-3} (m + 2)^{-3} \right] \\ + \frac{1}{27} \left[\frac{1}{2} + 4 \sum_{m=1}^{\infty} [m^{-3} (m + 3)^{-3}] \right] \\ + \frac{1}{64} \left[\frac{4}{27} + \frac{1}{32} + 4 \sum_{m=1}^{\infty} m^{-3} (m + 4)^{-3} \right] + \dots \text{etc.} \quad \dots\dots (64)$$

By expressing $1/[m^3(m + \alpha)^3]$ in partial fractions, we can derive the result

$$\sum_{m=1}^{\infty} \frac{1}{m^3(m + \alpha)^3} = \frac{6}{\alpha^5} \sum_1^{\infty} \left[\frac{1}{m} - \frac{1}{m + \alpha} \right] - \frac{3}{\alpha^4} \sum_1^{\infty} \left[\frac{1}{m^2} + \frac{1}{(m + \alpha)^2} \right] \\ + \frac{1}{\alpha^3} \sum_1^{\infty} \left[\frac{1}{m^3} - \frac{1}{(m + \alpha)^3} \right], \quad \dots\dots (65)$$

in which the three terms on the right may all be calculated easily for small values of α . The original expression (60) converges sufficiently rapidly for the sequence (64) to be cut off at, say, $\alpha = 4$.

The sum in the first term in (65) is simply

$$1 + \frac{1}{\alpha} + \dots + 1/\alpha,$$

while that in the second term is equal to the sum of the first α terms subtracted from $2 \sum_1^{\infty} 1/m^2$. In the third sum the terms cancel out with the exception of the first α terms which do not overlap. Values of the expression in (65) for $\alpha = 1, 2, 3, 4$ are given in table 3.

Table 3

α	1	2	3	4
$\sum_{m=1}^{\infty} m^{-3} (m + \alpha)^{-3}$	0.13040	0.03940	0.01687	0.00864

Substituting these values in (64) we obtain

$$S_1 = 0.816. \quad \dots\dots (66)$$

Referring to (26) it will be seen that $8S_1$ is small compared to $512C^3$ and therefore, in this instance, a more refined calculation of S_1 is unnecessary.

APPENDIX 2

In this appendix a closed expression is developed for the sum of the series

$$S_2(\beta) = \sum_{m=1}^{\infty} m^{-3} (2m + \beta)^{-3}, \quad \dots\dots (67) \\ (\beta = \pm 1, \pm 3, \dots).$$

Using the method of partial fractions we find

$$\sum_{m=1}^{\infty} \frac{1}{m^3(2m+\beta)^3} = \frac{24}{\beta^5} \sum_1^{\infty} \left[\frac{1}{m} - \frac{1}{m+\beta/2} \right] - \frac{6}{\beta^4} \sum_1^{\infty} \left[\frac{1}{m^2} + \frac{1}{(m+\beta/2)^2} \right] + \frac{1}{\beta^3} \sum_1^{\infty} \left[\frac{1}{m^3} - \frac{1}{(m+\beta/2)^3} \right]. \quad \dots\dots (68)$$

The three terms on the right will be considered in turn and it will be convenient to deal with positive and negative values of β separately. For the first term, with the notation

$$s(\beta) = \sum_{m=1}^{\infty} \left[\frac{1}{m} - \frac{1}{m+\beta/2} \right],$$

we have

$$\begin{aligned} s(\beta) - s(1) &= \lim_{M \rightarrow \infty} \left\{ \sum_1^M \left[\frac{1}{m} - \frac{1}{m+\beta/2} \right] - \sum_1^M \left[\frac{1}{m} - \frac{1}{m+\frac{1}{2}} \right] \right\} \\ &= \lim_{M \rightarrow \infty} \left\{ \sum_1^M \frac{1}{m+\frac{1}{2}} - \sum_{\frac{1}{2}(\beta+1)}^{M+\frac{1}{2}(\beta-1)} \frac{1}{n+\frac{1}{2}} \right\}, \end{aligned}$$

where

$$n = m + \frac{1}{2}(\beta - 1).$$

Thus, for $\beta \geq 1$,

$$s(\beta) - s(1) = \sum_1^{\frac{1}{2}(\beta-1)} \frac{1}{m+\frac{1}{2}} = 2 \left(\frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{\beta} \right).$$

Now
$$s(1) = 2 \left(\frac{1}{2} - \frac{1}{3} + \frac{1}{4} - \dots \right) = 2(1 - \log 2),$$

and hence
$$s(\beta) = 2 \left(\frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{\beta} - \log 2 \right). \quad \dots\dots (69)$$

For $\beta < -1$, we have

$$s(\beta) - s(1) = - \sum_{\frac{1}{2}(\beta+1)}^0 \frac{1}{n+\frac{1}{2}} = -2 \left(\frac{1}{1} - \frac{1}{1} - \frac{1}{3} - \dots + \frac{1}{\beta+2} \right),$$

and hence
$$s(\beta) = 2(1 - \log 2) - 2 \left(\frac{1}{1} - \frac{1}{1} - \frac{1}{3} - \dots - \frac{1}{\beta+2} \right). \quad \dots\dots (70)$$

The second term in (68) may be split up into two parts, of which the first is simply

$$\sum_{m=1}^{\infty} 1/m^2 = \zeta(2).$$

For the second part we use the relation

$$(1 - 2^{-x})\zeta(x) = 1^{-x} + 3^{-x} + 5^{-x} + \dots (x > 1), \quad \dots\dots (71)$$

which, for $x=2$, gives

$$\sum_1^{\infty} (m+\beta/2)^{-2} = 3\zeta(2) - \sum_1^{\frac{1}{2}(\beta+1)} (m-\frac{1}{2})^{-2}. \quad \dots\dots (72)$$

This holds for $\beta \geq 1$, while for $\beta \leq -1$ the corresponding expression is

$$3\zeta(2) + \sum_{\frac{1}{2}(\beta+1)}^{-1} (m+\frac{1}{2})^{-2}. \quad \dots\dots (73)$$

The third term in (68) is similar and we have, for the first part,

$$\sum_{m=1}^{\infty} 1/m^3 = \zeta(3),$$

while for the second part, using (71) with $x=3$, we find

$$\begin{aligned} &= 7\zeta(3) - \sum_{\frac{1}{2}(\beta+1)}^{1} (m - \frac{1}{2})^{-3}, \quad (\beta \geq 1), \quad \dots\dots(74) \\ \Sigma(m + \beta/2)^{-3} \} \\ &= 7\zeta(3) + \sum_{\frac{1}{2}(\beta+1)}^{-1} (m + \frac{1}{2})^{-3}, \quad (\beta \leq -1). \quad \dots\dots(75) \end{aligned}$$

From (68) the original sum for $\beta \geq 1$ is obtained from (69), (72) and (74):

$$\begin{aligned} S_2(\beta) = \frac{48}{\beta^3} \left(\frac{1}{1} + \frac{1}{3} + \frac{1}{5} + \dots + \frac{1}{\beta} - \log 2 \right) - \frac{6}{\beta^4} \left(4\zeta(2) - \sum_1^{\frac{1}{2}(\beta+1)} \frac{1}{(m - \frac{1}{2})^2} \right) \\ + \frac{1}{\beta^3} \left(\sum_1^{\frac{1}{2}(\beta+1)} \frac{1}{(m - \frac{1}{2})^3} - 6\zeta(3) \right). \quad \dots\dots(76) \end{aligned}$$

For $\beta \leq -1$ the corresponding expression is obtained from (70), (73), and (75):

$$\begin{aligned} S_2(\beta) = \frac{48}{\beta^5} \left[(1 - \log 2) - \left(\frac{1}{1} - \frac{1}{1} - \frac{1}{3} - \dots + \frac{1}{\beta+2} \right) \right] \\ - \frac{6}{\beta^4} \left(4\zeta(2) + \sum_{\frac{1}{2}(\beta+1)}^{-1} \frac{1}{(m + \frac{1}{2})^2} \right) \\ - \frac{1}{\beta^3} \left(\sum_{\frac{1}{2}(\beta+1)}^{-1} \frac{1}{(m + \frac{1}{2})^3} + 6\zeta(3) \right). \quad \dots\dots(77) \end{aligned}$$

For the special case $\beta=1$, (76) reduces to

$$S_2(+1) = 80 - 6[8 \log 2 + 4\zeta(2) + \zeta(3)],$$

and for $\beta = -1$, (77) gives

$$S_2(-1) = 6[8 \log 2 - 4\zeta(2) + \zeta(3)].$$

REFERENCES

- BLACKMAN, M., 1935, *Proc. Roy. Soc. A*, **148**, 365.
 BORN, M. and VON KÁRMÁN, TH., 1912, *Phys. Z.*, **13**, 297.
 BORN, M., 1923, *Atomtheorie des festen Zustandes* (Leipzig: Teubner).
 BROCH, E. K., 1937, *Proc. Camb. Phil. Soc.*, **33**, 485.
 DEBYE, P., 1912, *Ann. Phys., Lpz.*, **39**, 789.
 KNOPP, K., 1928, *Theory and Application of Infinite Series* (London: Blackie).
 THIRRING, H., 1913, *Phys. Z.*, **14**, 867.

The Interpretation and Application of Electron-Diffraction 'Kikuchi-Line' Patterns—Part I. The Determination of the Crystal Unit Cell, its Orientation and the Crystal Symmetry

By H. WILMAN

Applied Physical Chemistry Laboratories, Imperial College, London

MS. received 22 July 1947

ABSTRACT. The interpretation of electron-diffraction Kikuchi-line patterns from single crystals is developed as a powerful independent and general means of determining the crystal lattice type, dimensions and orientation and the Laue symmetry of the atomic arrangement. This aim is realized by a construction of the reciprocal lattice by methods which are described and illustrated. The satisfactory accuracy of the method is due to the relatively

precise definitions for the location of the Bragg reflection positions, derived and stated here for the first time, which enable the method to be used even when the crystal lattice constants are large so that the bands are narrow and diffuse-edged, as with rhombic sulphur. The patterns shown also illustrate features of the line intensities, especially the reduction in intensity when the bands pass near the undeflected-beam spot.

§1. INTRODUCTION

SINGLE-CRYSTAL diffraction patterns provide much more complete though usually also more complex information than is yielded by the ring patterns obtained from polycrystalline material, with respect to the structure of the material, i.e. lattice type, symmetry and atomic arrangement. It is therefore especially important to be able to interpret and utilize single-crystal patterns in such a way as to reveal all the information available in a form easy of application to all cases. The interpretation of the single-crystal Kikuchi-line pattern is developed below as a powerful independent technique by which the structure and orientation of single crystals can be determined.

Electron-diffraction patterns of black and white lines from a single crystal of mica were first observed in 1928 by Kikuchi, who showed the relationship between the line positions and the crystal net planes, and the rough agreement between the relative line intensities and the relative x-ray reflecting powers of the orders of reflection from a series of net planes. Since then, although such patterns have been used as a means of defining the orientation of crystals of known structure (e.g. Thirsk and Whitmore 1940) and of checking the structures of graphite and cadmium iodide (Finch and Wilman 1936, 1937 a) they have otherwise only been made use of in a general way, by inferring from their presence that there are relatively large thicknesses (several hundred Å. or more) of coherently diffracting crystals in the specimen, and from their diffuseness or displacement the degree of fine-scale surface roughness, mosaic structure and distortion of the crystal.

In order to use these patterns to determine the lattice structure, orientation and Laue symmetry of a crystal, all that is required is to construct the reciprocal lattice, and this can be achieved in the way described below. Moreover, it can be done independently of whether or not a clear pattern of Laue spots is also present in the pattern, and independently of any other source of knowledge. The crystal lattice and its orientation can be determined from a single photograph if this records a wide enough range of diffracted rays, i.e. the usual 10 to 20° from the primary beam direction.

The method is also of general applicability, for it has been demonstrated (Finch, Quarrell and Wilman 1935, Finch and Wilman 1936, 1937 a, 1937 b) that clear Kikuchi-line patterns can be obtained by transmission through thin cleavage flakes or natural lamellar crystals in favourable cases, or by reflection from (a) smooth natural faces, (b) cleavage faces, (c) 'parting' fractures, (d) conchoidal fractures, (e) polished crystal surfaces (plane or curved), or (f) polished and lightly etched surfaces. It is the only stationary-crystal electron-diffraction method applicable if the patterns contain few or no Bragg or Laue spots owing to the narrowness of the Laue zones. In such cases the crystal surface can usually be slightly roughened by careful abrasion or etching (see examples of FeS₂ and ZnS respectively: Finch and Wilman 1937 b) so as to yield more extensive spot patterns, or the crystal can be rotated during exposure of the plate, but the accuracy

of location of the lattice rows from the spacings and curvatures of the spot rows, i.e. the Laue zones, is in any case often very much less than in the Kikuchi-line method, in which angles between net-planes are directly measurable.

§ 2. THE BASIS OF THE INTERPRETATION OF KIKUCHI-LINE PATTERNS

(i) *The origin and general nature of the lines and bands*

For the present purpose it suffices to apply, as Kikuchi did, Rutherford and Andrade's (1914) basic conception of the line pairs as originating by selective Bragg reflection by the crystal net-planes from the diffuse background scattering, which has a radial fall of intensity with increasing angle of deviation from the primary beam. A practically parallel pair of black and white lines in the pattern is thus the intersection made with the photographic plate by the two halves of a cone round the net-plane normal, with semi-apical angle $90^\circ - \theta$; θ is the Bragg angle, related to the net-plane spacing d and the electron wavelength λ by Bragg's law, $2d \sin \theta = n\lambda$, where n is an integer, the order number of the diffraction. It is found (Kikuchi 1928, Shinohara 1932, Kirchner 1932, Emslie 1934, Finch, Quarrell and Wilman 1935, Finch and Wilman 1936, 1937 a, 1939) that the measured line-pair separations correspond well (to within about 1 to 2%) with this simplified interpretation, the wavelength of the electrons which contribute to the lines being practically identical with that of the primary beam.

In cases so far examined (Kikuchi 1928, Kirchner 1932) the relative intensities of the different line-pair diffractions also show a general correspondence with the relative structure factors and electron or x-ray reflecting powers of the net-planes when allowance is made for the different inclinations of the planes to the beam. The relative intensities of orders of reflection from a net-plane are, however, appreciably modified by secondary scattering, or more correctly by dynamic interaction of the scattered wavelets, so as to even them out partially. In certain definable cases orders of reflection may appear which are forbidden according to the theory of single-scattering, but this cannot affect extinctions characteristic of the lattice type (Brandenberger 1937).

Kirchner (1932) has pointed out that the elementary diffraction theory for single scattering also leads to a Kikuchi-line breadth proportional to the corresponding plane spacing d/n , and inversely proportional to the thickness of the diffracting part of the crystal in a direction normal to the net-plane. The lines corresponding to d/n about 2 Å. or more are in fact generally very broad and diffuse, while those of higher orders from the same plane or from planes with lower d/n are progressively much sharper. The dynamical theory of scattering has been found by Lamla (1938) to give an expression for the line breadth proportional to $1/\sin 2\theta$, which is nearly the same as Kirchner's result, since 2θ is small for fast electrons. Though the theory for selective single scattering from the diffuse background goes so far to account for the position and form of the lines it fails to explain the existence of the prominent bands with symmetrical intensity distribution, which still appear, except near the primary beam, even when the net-plane is parallel to the primary beam. Hayasi (1934), however, showed that such bands are to be expected if from each scattering centre spherical waves spread out with amplitude proportional to the cosine of the angle of deviation from the primary beam, and these are scattered again once by all the other atoms in the crystal. The more complete dynamical theory, applied by Shinohara (1932 a), also yielded an

intensity distribution across the symmetrical bands which was similar to that observed (figure 1(b)). For present purposes, however, these details of the intensity distribution need not be considered further.

(ii) *Location of the Bragg reflection positions and measurement of their separation D*

For present purposes it is first desirable to define more accurately than has hitherto been done the positions of the Bragg reflection directions relative to the Kikuchi lines and band edges, which are very diffuse when they arise from planes whose spacings are large.

In the case of the symmetrical bands associated with planes that are parallel to the primary beam it appears to have been generally assumed that the Bragg positions are at the band edges; these, as the author now defines them, are the points where the intensity is zero relative to the general background in the region considered, as it passes from a positive value (excess over the background) near the maximum on the side within the band, to a negative value which reaches a minimum of intensity slightly farther out. For example, Thomson and Cochrane (1939, pp. 110-116) seem to indicate this to be their opinion, and in indexed patterns of Emslie (1934), Tillman (1935), Finch, Quarrell and Wilman (1935) and Finch and Wilman (1936, 1937 a), the lines denoting the consecutive orders of Bragg reflection from such planes are drawn in accordance with this view. No close quantitative test yet appears to have been made, and this is now supplied by measurements on patterns taken at not too small a camera-length (~ 22 cm.) and obtained with a finely-focused narrow electron beam (~ 55 kv.) reflected from a perfect magnesium-oxide cleavage face (cube face (001)), so that the lines are as sharp as possible.

Figure 1(a) shows part of such a pattern where the median of the strong vertical band system due to the (100) plane lies in the plane of incidence, and the undeflected-beam spot lies on this median 3.8 cm. below the lower edge of the figure. The intensity distribution across this band system in a region where obliquely crossing bands are not prominent, is of the form shown in figure 1(b). In this figure are also shown measurements made on the plate at the 400, 600 and 800 band edges, corresponding to points where the intensity was estimated visually to be the same as that of the general background in the same neighbourhood.

It will be seen that there is very good agreement in the spacings for the strong and relatively sharply-defined 400 and 600 band edges, and only slightly less so for the faint but sharp 800. The positions for the very diffuse 200 band edges calculated from these measurements were found also to agree excellently with the points where the intensity was the same as that of the background. These regularly spaced positions are thus evidently the successive orders of the Bragg reflection positions. It is also clearly shown in figure 1(a) that the vertically elongated but laterally very sharp Laue spots lie in every case within 0.001 cm. of these positions when the net-plane passes exactly through the primary beam. (Actually, in figure 1(a) the Kikuchi line pattern is not exactly symmetrical about the plane of incidence, but is displaced sideways by 0.007 cm. relative to the Laue-spot rows, due to an azimuth deviation of $0^\circ 1'$ from the symmetrical one, but the quantitative correspondence is clear.) This fact confirms that these are the Bragg reflection positions for electrons which have practically the same speed and energy as those of the incident beam; in the case of the very sharp higher-order band edges such as 800 and 10, 0, 0 the electrons

contributing to the Kikuchi line pattern are calculated to have lost not more than 50 electron-volts energy.

In the case of net-planes not parallel to the primary beam, black and white line-pairs arise in general, but here also there is a band of increased intensity

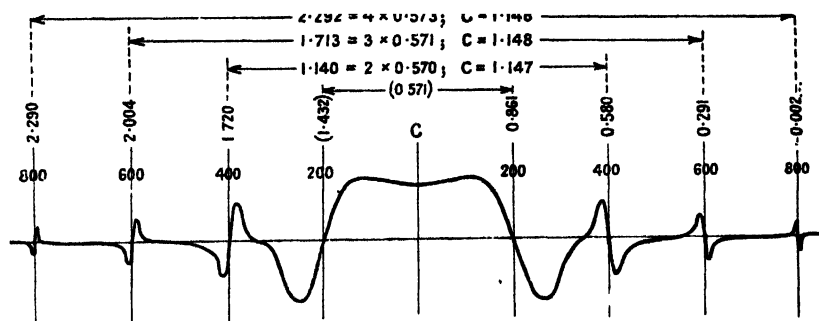


Figure 1 (b).

between the lines, which is strong when the line-pair separation is small, corresponding to planes of high spacing. This can be seen prominently in figure 2(a), which shows also the sector-like effect caused by the crossing and superposition of parts of two or more such bands. In these unsymmetrical bands and line-pairs it seems hitherto to have been usual to measure the distance between the centres of the black and white lines, assuming these to occur at the Bragg reflection positions.

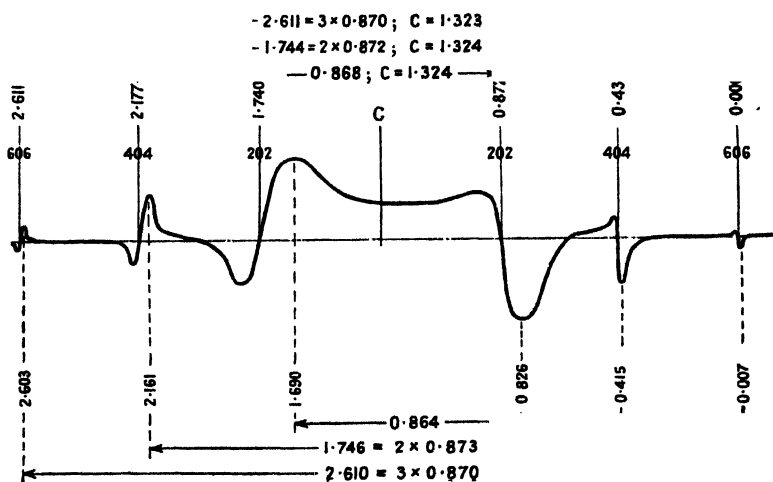


Figure 2 (b).

To test this, measurements have now been made on several successive orders of line-pairs or band edges such as those in figure 2(a), which is also part of a pattern from an MgO cleavage face at an azimuth near that of figure 1(a); thus the pattern adjoins on the right that recorded in figure 1(a). An example of the measurements, for the orders of diffraction from the (101) plane, is given in figure 2(b), at the points indicated relative to the intensity distribution shown qualitatively in the figure. It will be seen from the upper series of measurements that the points

plane through O parallel to a net-plane system (hkl) in the crystal, inclined at ϕ to OO_1 . Then the radius vector r^* from O to the reciprocal-lattice point $[[hkl]]^*$ is $h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$ and is along the normal OP' to this plane, with length $1/d_{hkl}$. To define the position of this point $[[hkl]]^*$ and the symmetrical point $[[\bar{h}, \bar{k}, \bar{l}]]^*$ in space, it is only necessary to specify the direction and the magnitude $r^* \cos \phi$ of the projection of r^* on the plane through O normal to OO_1 (parallel to the plate), together with the distance $r^* \sin \phi$ of the point from this plane, which is considered positive if measured from the plane along OO_1 and negative if along O_1O .

These magnitudes can be obtained easily from the camera-length L and measurements of the perpendicular p from O_1 to the line-pair median, and of the separation D of the line-pair K_1K_2 , measured along the perpendicular O_1P . Thus

$$r^* = 1/(d/n) = \sin \theta / \lambda \simeq \theta / \lambda \simeq (D' \cos \phi) / \lambda L \simeq (D \cos^2 \phi) / \lambda L.$$

$$\cos \phi = L / (L^2 + p^2)^{1/2}; \quad \sin \phi = p / (L^2 + p^2)^{1/2}; \quad \tan \phi = p / L.$$

Whence

$$r_p^* = r^* \cos \phi \simeq (D \cos^3 \phi) / \lambda L$$

$$= \left(\frac{D}{\lambda L} \right) \cdot \left\{ 1 - \frac{3}{2} \left(\frac{p}{L} \right)^2 + \frac{15}{8} \left(\frac{p}{L} \right)^4 - \frac{35}{16} \left(\frac{p}{L} \right)^6 + \dots \right\}$$

$$= (1/\lambda L) \cdot D \cdot f(p/L),$$

$$z = r^* \sin \phi = (p/L) \cdot r^* \cos \phi.$$

Figure 4(a) can be used to find $f(p/L)$ for any values of p and L likely to be encountered normally, since under the horizontal p/L scale are marked scales of p for a range of L values. The projection of the reciprocal lattice is drawn by plotting as radius vectors $\lambda L \cdot r^* \cos \phi$, i.e. $D \cdot f(p/L)$, to a convenient scale, for example a five (or sometimes ten) times scale, the corresponding distances $z \cdot \lambda L$ of the reciprocal-lattice points from the projection plane $(p/L) \cdot D \cdot f(p/L)$, being written by each point as the actual distance in centimetres relative to this same scale.

An exact expression for D , i.e. $PK_1 + PK_2$, is obtained from triangles OPK_1 and OPK_2 in figure 5:

$$PK_2 = L' \sin \theta / \cos(\phi + \theta); \quad PK_1 = L' \sin \theta / \cos(\phi - \theta),$$

thus

$$D = PK_1 + PK_2 = L' \sin 2\theta \cos \phi / (\cos^2 \phi - \sin^2 \theta)$$

$$\sim L' \sin 2\theta / \cos \phi \simeq 2L' \sin \theta / \cos \phi$$

$$= 2L \sin \theta / \cos^2 \phi.$$

The last approximate expression for D is accurate to within 0.5% even when ϕ is as much as 20° ($p \sim 7.7$ cm. when $L = 22$ cm.) and θ as much as 4° ($D \sim 3$ cm. with $L = 22$ cm.), thus the same accuracy, sufficient for present purposes, is obtained in $r^* \cos \phi$, because from this D value we have without further approximation:

$$r^* \cos \phi = (n/d) \cos \phi = (2 \sin \theta \cos \phi) / \lambda = (D \cos^2 \phi) / \lambda L' = (D \cos^3 \phi) / \lambda L,$$

which is the expression already given above.

Similarly, to find how far the trace of the net-plane on the plate is displaced from the actual line-pair 'median', midway between K_1 and K_2 and normal to O_1P , we have

$$(PK_2 - PK_1) / D = L' \sin \theta \cdot \{1/\cos(\phi + \theta) - 1/\cos(\phi - \theta)\} / D$$

$$= 2L' \sin^2 \theta \cdot \sin \phi / (\cos^2 \phi - \sin^2 \theta) D$$

$$= \tan \theta \tan \phi$$

$$= (p/L) \tan \theta \simeq D_p / 2L(L + p^2)^{1/2},$$

by using the above accurate expression for $PK_1 + PK_2 = D$. Since θ is never likely to exceed 6° in normal practical cases, it follows that even if p/L is as high as 0.3 and $\theta = 6^\circ$, the plane trace is displaced from the 'median' by only $(0.3 \times 0.1)D/2$ i.e. $0.03(D/2)$, which can be neglected for present purposes.

A further consideration sometimes required is the variation of line-pair separation D_y with distance y along the 'median' or net-plane trace, since the distance apart D , i.e. K_1K_2 , at the hyperbolae vertices ($y=0$) is sometimes not directly measurable because this region lies outside the recorded photograph. It can be shown that

$$D_y^2 = D^2 + 4y^2(\cos^2 \phi \cot^2 \theta - \sin^2 \phi),$$

whence

$$(D/D_y)^2 = 1 - (4y^2/D_y^2) \sin^2 \theta (\cos^2 \phi - \sin^2 \theta).$$

As shown above,

$$D \sim 2L \sin \theta / \cos^2 \phi,$$

thus, neglecting $\sin^2 \theta$ in comparison with $\cos^2 \phi$ we have, when $D_y \sim D$,

$$(D/D_y)^2 \sim 1 - (y^2/L^2) \cos^2 \phi = 1 - (y^2/L^2) \{L^2/(L^2 + p^2)\}.$$

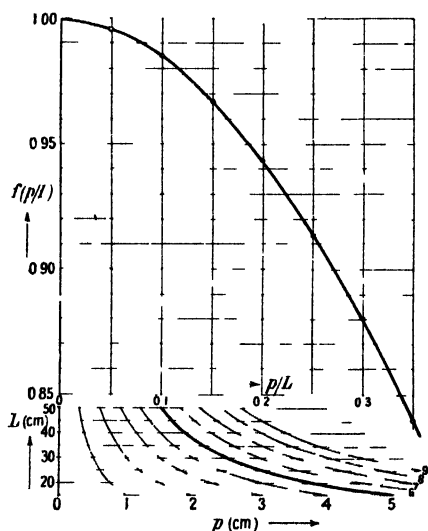


Figure 4(a).

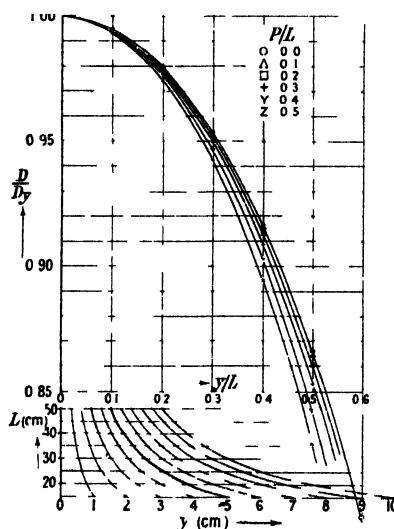


Figure 4(b)

Figure 4(b) shows the values of D/D_y corresponding to this expression, over the range of y/L and p/L concerned, from which can be interpolated in any usual case the factor D/D_y by which D_y must be multiplied to estimate D .

(ii) *The relation between a 'positive' print and the initial 'negative' recording*

It is convenient to measure D and p of the line-pairs on either a 'positive' or a 'negative' reproduction of the initial photographic 'negative' recording of the electron diffraction pattern.

If a negative reproduction is made by first printing a positive on to a plate and taking a print from this on paper, it can be considered in its geometry as identical with the initial negative recording.

A positive reproduction, however, is usually preferred owing to its greater clearness of detail, and it bears in its geometry the relation of a mirror image to the

initial negative. More exactly, if the plate and print are viewed face side up, with top edges roughly in line, and then by rotation of one a line BA in the print is placed in line with the corresponding line AB in the plate, then the print is the mirror image of the plate (apart from interchange of black and white) in a plane which is normal to the line chosen, and cuts it midway between the pair of 'A' or 'B' points.

Alternatively, if the positive print and the plate are put in register with the sensitive plane faces together and are then opened apart by rotating the print about *any* line in their plane until the sensitive faces are again coplanar, then the print and plate are mirror images of each other relative to that line. It follows that, in respect of its geometry the 'positive' print can be regarded as identical with the initial 'negative' which would have been obtained if a crystal had been used which was the mirror image of the crystal actually used, in respect of its lattice, orientation and atomic structure. If the crystal lattice and the atomic structure both possess a plane of symmetry (or have a Laue symmetry containing one) then the details of the positive print are congruent with those of an initial 'negative' recording obtainable from an *actual* crystal if it were in an orientation which is the mirror image of the actual orientation in a plane perpendicular to that of the negative and its positive print (both with sensitive sides up), and passing through the line about which the print was rotated to being it facing the same way up as the negative.

It is this equivalence which has been invoked hitherto in most cases where positive prints of photographs from a crystal known to have a plane of symmetry have been indexed, both Kikuchi-line and spot patterns having been indexed as if they were the initial 'negative' recording (apart from interchange of black and white) viewed along the beam direction from behind the crystal (e.g. in Finch, Quarrell and Wilman 1935, and Finch and Wilman 1935, 1937 a, 1937 b).

However, it is preferable in such cases, and it is necessary in the case of crystals of unknown structure or those known to have no plane of symmetry, to adopt the following unambiguous system which will be applied in the examples given below. The pattern, whether a positive or a negative, can be thought of as situated so that the features of its intensity distribution coincide with the corresponding distribution of the electrons in the recording plane; a 'negative' print is then to be considered as if it were the actual negative recording viewed along the primary beam direction, from behind the crystal (figure 3), and a 'positive' print is to be regarded (apart from interchange of black and white) as if it were the initial negative recording viewed from behind, in the direction opposite to that of the primary beam, i.e. looking towards the crystal.

(iii) *Practical details of pattern measurement and projection construction*

Location of O_1 . Though it is not absolutely essential to locate O_1 , the foot of the perpendicular from the crystal to the plate, to within more than about 2 cm. when L is about 20 to 25 cm., it is obvious that the more accurately this is done the more consistent will be the estimated lattice-point positions and the accuracy of determination of the lattice symmetry and dimensions. One indirect way of locating O_1 sufficiently closely is from observations of the positions of the focused undeflected-beam spot F (figure 5), recorded in the photograph, and of the unfocused-beam spot U, the camera-length L , and the distances BA and BC of the focusing-coil from the anode diaphragm A and the plate. Thus, approximately,

by assuming that any change in beam direction caused by the focusing-coil occurs near the central plane of the coil, we have

$$\begin{aligned} CO_1 &\approx BE + E_1O_1 \\ &\approx BE + (CF - BE) \cdot \{1 - (OO_1/BC)\} \\ &= BE \cdot OO_1/BC + CF \{1 - (OO_1/BC)\} \\ &= (OO_1/BC)(AB/AC) \cdot CU + \{1 - (OO_1/BC)\} \cdot CF. \end{aligned}$$

AC, AB, BC, CU and CF can all be measured, and the vectorial form of the expression applies to both the cases (a) and (b) of figure 5, though that of (a) gives the best beam focusing, making F as near U as possible, and it should therefore always be used. If F and U coincide,

$$CO_1 = CF \cdot (AC' - OO_1)/AC.$$

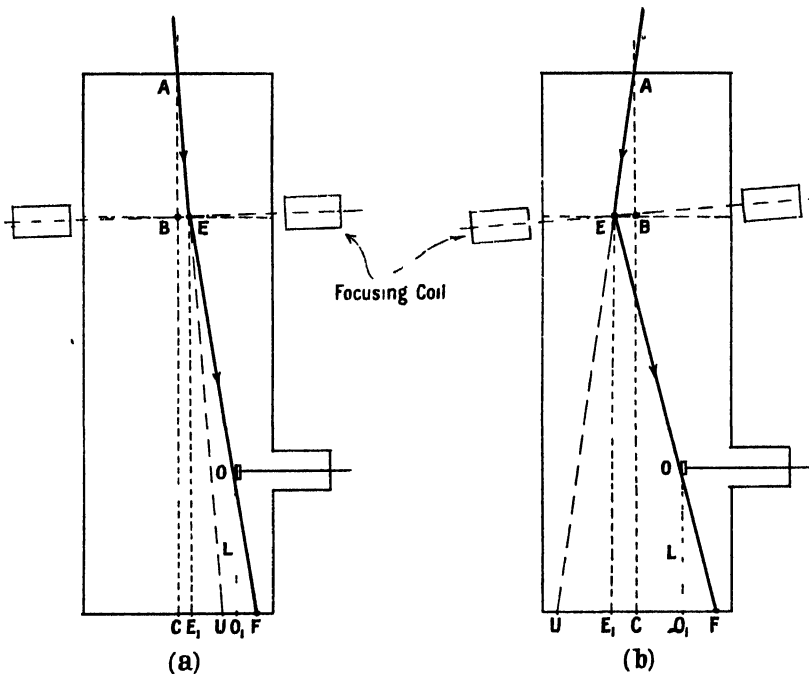


Figure 5.

Measurement of D and p . When a print of the negative has been prepared and stuck down on a suitable-sized piece of paper, the line pairs are sorted out and extensions drawn round the print so as to label each of the line-pair medians with a serial number, as in figures 7(a) and 8(a). Corresponding black and white lines are practically parallel and of similar intensity and sharpness. The fainter higher orders are not always visible continuously across the whole photograph, but usually appear more clearly on regions of raised intensity due to other band systems. The line-pair separation D at or near the foot of the perpendicular p from O_1 to the lines are then measured on the print, using a thin-edged finely-graduated millimetre scale to avoid parallax, and measuring D usually to an accuracy of 0.1 mm. with the aid of a hand lens magnifying about five times. The perpendicular distances p of the medians from O_1 are also

measured, usually to 0.5 or 0.3 mm., or if possible, to 0.1 mm. in cases where p is only a few millimetres; and a table of D and p values is constructed from which $D \cdot f(p/L)$ and $(p/L)D \cdot f(p/L)$ values can be calculated.

Where possible D should be derived from measurements between the sharper higher orders, divided by the order number. If it is necessary to measure between faint lines, the average of several estimates should be taken. Note especially that when the main line pairs have been sorted out the sharpness of any intensified line can be used to give an indirect estimate of D and so usually leads to identification of the practically parallel absorption line (with the help of a lens), even when this is too faint to be noticed immediately.

If D cannot be measured directly because the region near the foot of the perpendicular from O_1 to the median lies outside the recorded photograph, the spacing D_{\parallel} must be measured where possible at a distance y from this region, and D estimated by use of figure 4(b), as described in §3(i).

Construction of the reciprocal-lattice projection. It is desirable to mark points on both sides of the origin, which is chosen some 20 cm. or more to one side of the print, and the geometrical positions of several higher orders should be marked even if they are too faint to be clearly distinguished in the pattern. The serial number of the net-plane is marked below each order, with the order number as suffix if desired. Accuracy of location of the radius-vector normals to the line-pair medians is essential and is assisted by careful use of transparent 60° set-squares of 10-inch hypotenuse. Allowance has to be made for the appreciable opposite curvature of the components of a line-pair in estimating the direction of the median, by finding the direction midway between chords or tangents in the same region of the line-pair.

The final stage is to write above each point in the projection the calculated distance of the point from the projection plane, with the appropriate sign. The projection is to be considered always as viewed from below the recording plane, i.e. in a direction opposite to the primary beam, so that the positive z distances (along OO_1 , as defined in §3(i)) are from the projection plane towards the observer, and negative distances are on the opposite side. With the sign convention stated above and if a positive print is used, the projection points whose radius vectors are in the same direction as that in which p is measured from O_1 will have negative distances from the projection plane (cf. figure 3).

For the purpose of determining the lattice shape, dimensions and orientation, it is desirable to mark the reciprocal-lattice points as small points surrounded by a ring, but it will often be useful to make copies showing the points as black discs with diameters a measure of the intensities of the diffractions.

(iv) *Use of scales to read off $\lambda L \cdot r_p^*$ and λLz directly from line positions*

When the line-pair separations D are mostly between 1 and 2 cm., and the patterns not too intricately criss-crossed, time and trouble can be saved with but little loss in accuracy by the use of measuring scales from which $D \cdot f(p/L)$, i.e. $\lambda L \cdot r_p^*$ and $D(p/L) \cdot f(p/L)$, i.e. λLz can be read off directly as the difference in the scale readings at the black-and-white-line Bragg positions when the scale is placed normal to the lines with its zero at O_1 .

The possibility of constructing such scales is due to the fact that the factor $f(p/L)$, by which D is multiplied to obtain $\lambda L \cdot r_p^*$, is not far from unity in all cases.

likely to be needed, and that its variation is small over a range of p/L of approximately 0.1 corresponding to band widths D up to 2 cm. when $L \approx 20$ cm. Under these conditions if a function, y , of p is derived such that

$$dy/dp = f(p/L),$$

we can take as a close approximation

$$\Delta y = \Delta p \cdot f(p/L), \quad \text{or} \quad (y_2 - y_1) = (p_2 - p_1) \cdot f\{(p_1 + p_2)/2L\}$$

even if $p_2 - p_1$ is made so large as to take the value of the line-pair separation D , so that

$$y_2 - y_1 = D \cdot f\{(p_1 + p_2)/2L\}.$$

Correspondingly, to construct the scale to read off $\lambda L \cdot z$ directly, we require a function y' given by

$$dy'/dp = (p/L) \cdot f(p/L).$$

We have thus:

$$f(p/L) = \cos^3 \phi = [1 + (p^2/L^2)]^{-3/2},$$

$$y = \int f(p/L) dp = p \cdot f_2(p/L) = p \cdot p \{1 - f_2(p/L)\},$$

$$y' = \int (p/L) \cdot f(p/L) dp = L \{1 - f_2(p/L)\}.$$

The values of $f_2(p/L)$ calculated are given in table 1, whence values of $(p - y)$ and y' were calculated corresponding to a range of p values from $L = 15, 20, 25, 30, 35, 40, 45$ and 50 cm. From these data scales were drawn which were then photographically reduced to the required size. A typical scale is shown in figure 6. Similar scales for any required intermediate value of L can be constructed if necessary in the same way.

Table 1. Variation of $1 - f_2(p/L)$ with p/L

p/L	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5
$1 - f_2(p/L)$	0.00125	0.0050	0.01106	0.0194	0.0299	0.0422	0.0561 ₅	0.0715	0.0880	0.1054

A trial of such scales over a wide range of p and D showed excellent agreement with the calculated values of $D \cdot f(p/L)$ and $(p/L)D \cdot f(p/L)$. The best scale for practical use is a reproduction on a photographic plate, but it is also possible to use a paper print of the appropriate scale, stuck on the underside of a piece of transparent plastic sheet, if the shrinkage of the paper after processing is found to be negligible. The zero is placed at O_1 then the scale moved slightly if necessary so that a scale mark comes opposite one Bragg position, whence only a single interpolation to a tenth of a division is necessary at the other Bragg position and the difference of the two readings is easily found and is then written down in the list of pattern data. When the list is completed the projection can then be constructed as described above.

§ 4. CHOICE OF AXES, DETERMINATION OF SYMMETRY AND ORIENTATION, INDEXING OF PATTERN, AND CALCULATION OF LATTICE CONSTANTS

When an adequate number of points are located in the projection and the distances of the lattice points from the projection plane are marked, it is easy to recognize the main reciprocal-lattice rows, because any straight line in space is also represented by a straight line in the projection and, further, the distance λLz of the lattice points from the projection plane must be proportional to their distance along

the line in the projection. The rows with densest lattice-point population are of course the most prominent. Similarly, it is easy to recognize which are the main lattice planes because they contain two or more systems of parallel densely-populated lattice rows. As a convenient 'scaffolding' or reference of coordinates to assist in this process, it is helpful to draw lightly the main lattice lines through the coplanar set of points (forming a two-dimensional lattice) due to the planes of a main zone, usually one whose zone axis is near the primary beam direction, so that the points in the reciprocal lattice plane correspond to the Laue spot pattern in the circular spot zones.

Having tentatively drawn in pencil some coplanar sets of main lattice rows in a favourable part of the projection, so as to show part of the cross-grating set of

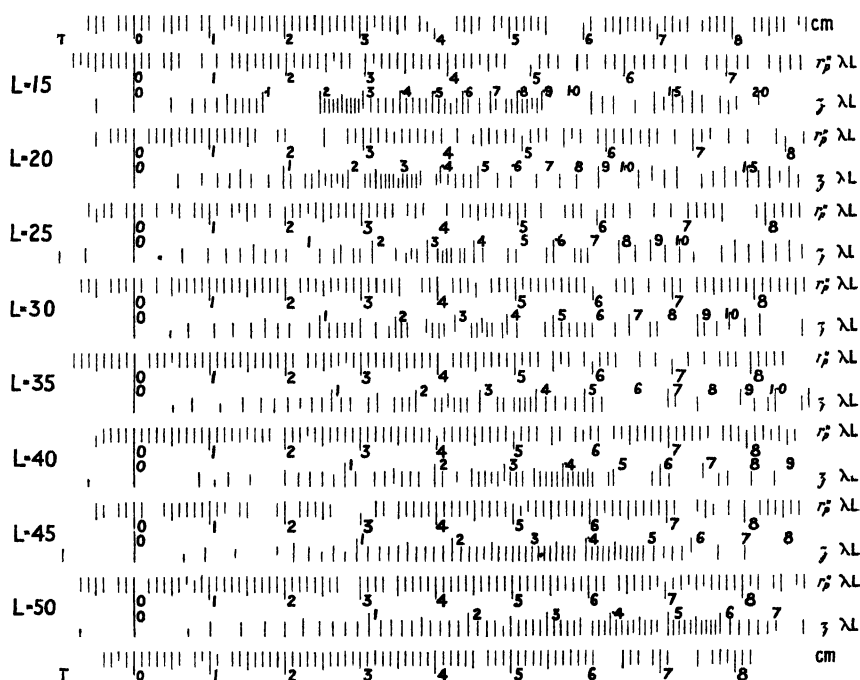


Figure 6

lattice points in such planes, a different coplanar set can be explored and delineated in the same or another part of the diagram. From such initial trials it is usually clear whether angles of 90° or 120° occur, and which is the appropriate crystallographic system of axes and lattice symmetry, and the whole diagram is then mapped out in lattice rows and planes accordingly, so as to outline the neighbouring unit cells of the reciprocal lattice; the indexing of the lattice points then follows easily as the simple process of writing down the coordinates of the points relative to the chosen axes.

From a consideration of the indices of the observed diffractions, supplemented by similar data from photographs obtained at other crystal settings, any systematic extinction of definite types of diffractions can be recognized and used in the usual way to determine the possible space groups characteristic of the atomic arrangement in the crystal. Allowance must be made, as in x-ray diffraction, for the fact that the "Laue symmetry" shown by diffraction patterns includes a centre of

symmetry even if there is none in the symmetry of the crystal structure; and other data such as observation of crystal forms and etch figures is required to decide this point. In electron diffraction, as already noted, there is also the possibility in certain cases that orders of reflection from a net-plane may occur with appreciable intensity, due to secondary or dynamic scattering, which would otherwise be forbidden according to the more approximate single-scattering theory. The further consideration of Kikuchi-line intensities will be left to a later contribution.

The angles between the reciprocal-lattice rows in planes inclined to the main projection plane, and the spacings between the points in the rows, can be estimated by the usual methods of projective geometry. Side projections in suitable directions, on planes perpendicular to the initial projection plane are especially useful

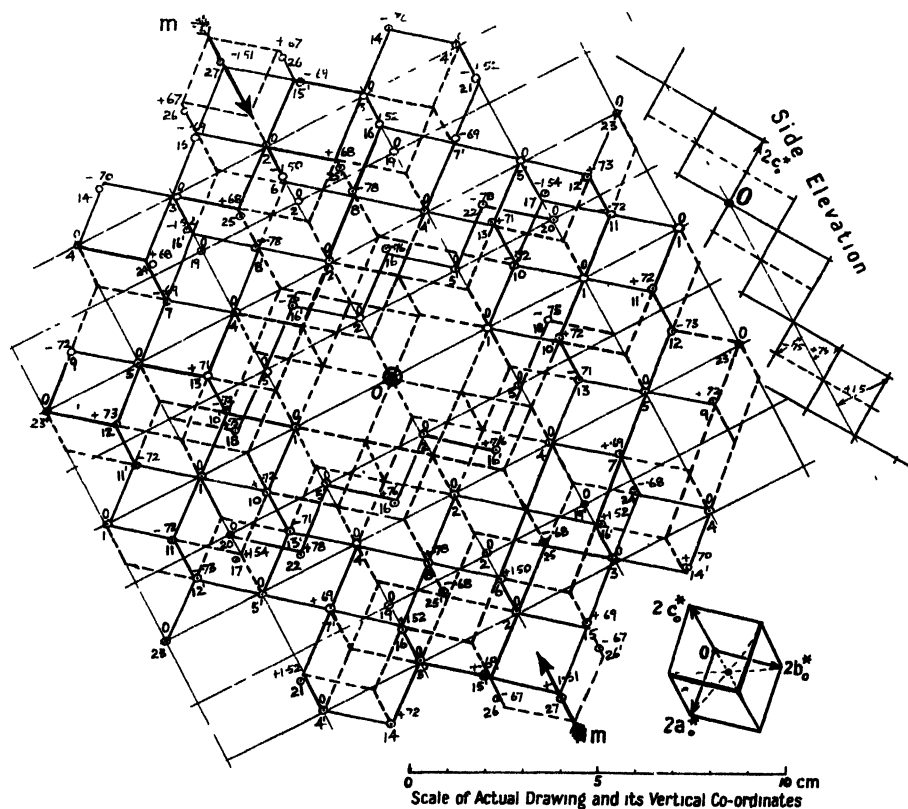
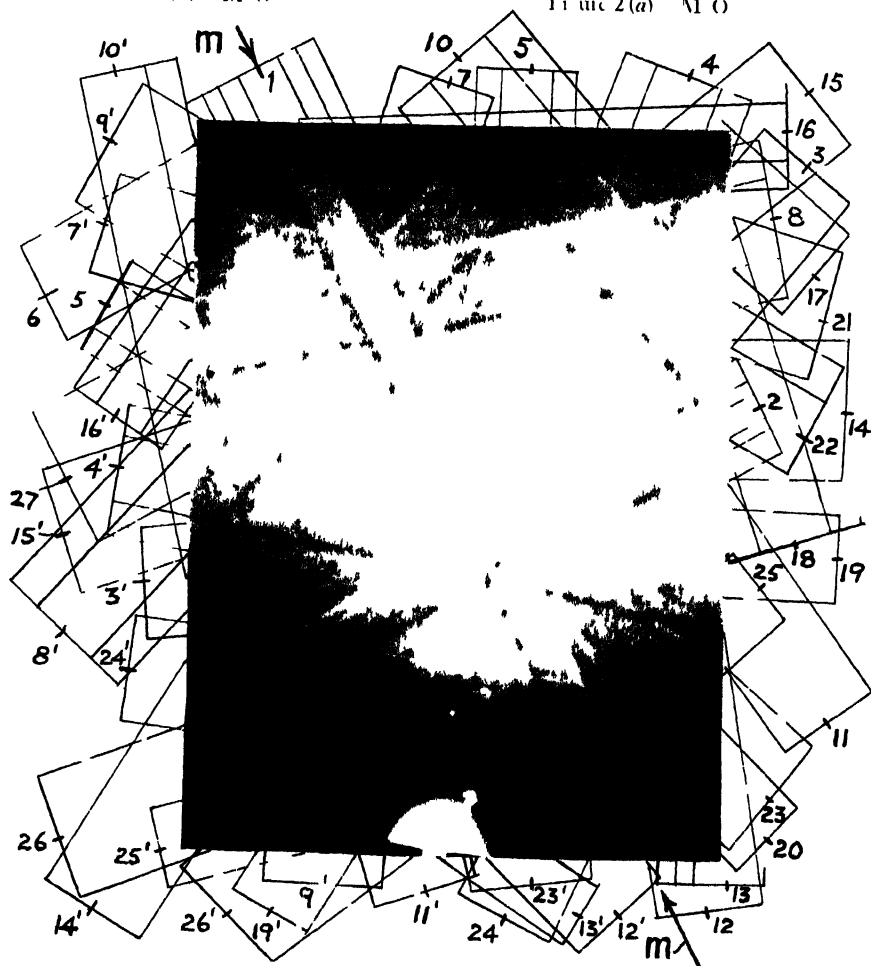
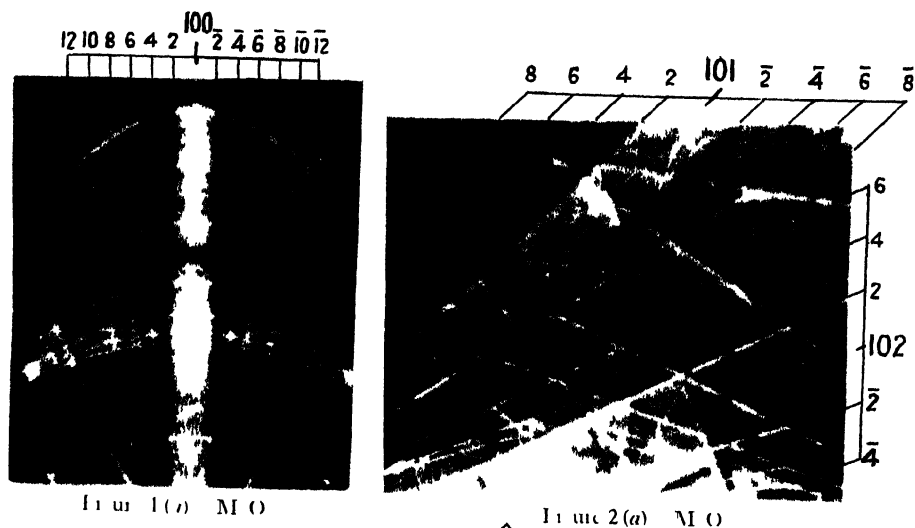


Figure 7 (b).

when these can be drawn so as to give desired angles and spacings directly (see figures 7(b) and 8(b)). In awkward cases it may sometimes be most direct to construct the lattice-point positions in space by mounting the projection on cardboard and sticking pins through the projected points so that their heads are at the appropriate heights. For convenience all the points can be brought on to one side of the cardboard base by adding a constant height to the calculated distances of the points from the projection plane.

The reciprocal-lattice angles and the lengths proportional to its axes are estimated from the projection or the model by taking averaged directions and lengths most consistent with the constructed lattice point positions over a region comprising as many unit cells as possible; the actual reciprocal axes are then



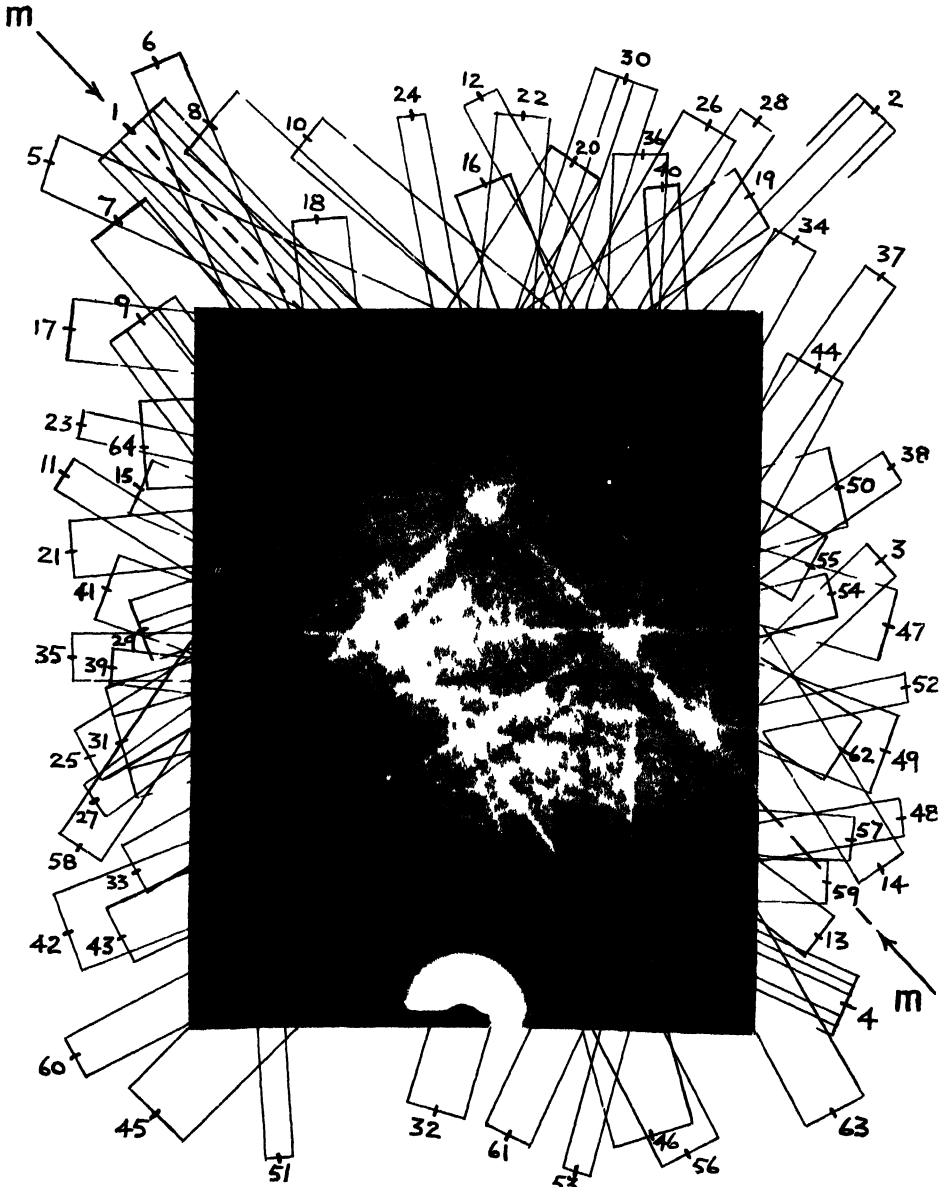


Figure 8(a) Sulphur conchoidal fracture. $I \sim 22 \text{ cm}$, 54 kV. O_1 at intersection of 4 and 37 band mediums

calculated by dividing the above lengths by the scale factor of the drawing or model, and by the estimated λL for the photograph. From the a^* , b^* , c^* , α^* , β^* , γ^* so obtained, the crystal lattice constants a , b , c , α , β , γ can be calculated from the usual relationships. The symmetry of the crystal lattice is of course of the same kind as that of the reciprocal lattice.

Although the determination of the lengths a , b , c involves the value of λL , which is not usually known with great accuracy, the axial ratios and the angles α , β , γ can be obtained fairly accurately as above from the reciprocal lattice measurements, or also from the angles between net-planes by the usual methods used with

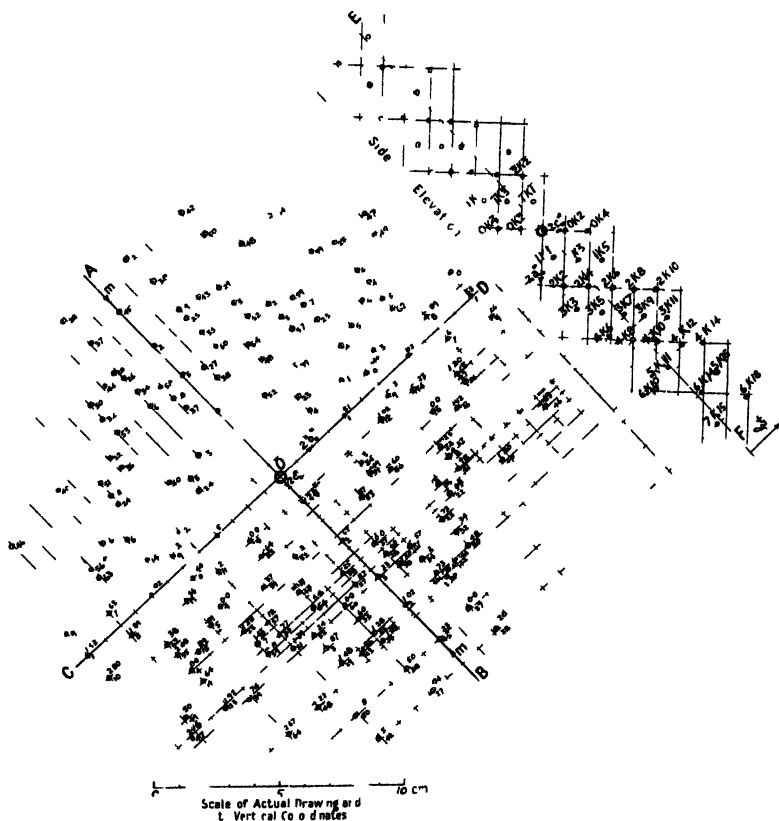


Figure 8(b).

goniometric measurements obtained from crystal faces. In any case, measurement of the angles $(180^\circ - \delta_p)$ enclosing O_1 , between suitable line-pair medians in the pattern, can be used to calculate the angles δ between the corresponding net-planes, as check comparisons with the δ values calculated from the lattice constants obtained from the reciprocal-lattice projection. The angles $(180^\circ - \delta_p)$ can usually be estimated to within about $0^\circ 10'$ by use of a normal set-square and a transparent rotating-arm set-square having a 4-inch radius scale of degrees. To obtain δ from the measured angle $(180^\circ - \delta_p)$ and the p_1 and p_2 values for the two line-pair medians, we have simply

$$\cos \delta_p = \{\cos \delta - \cos(90^\circ - \phi_1) \cos(90^\circ - \phi_2)\} / \sin(90^\circ - \phi_1) \sin(90^\circ - \phi_2),$$

$$\text{so that} \quad \cos \delta = \{\cos \delta_p + (p_1 p_2 / L^2)\} \cdot \{1 + (p_1 / L)^2\}^{-\frac{1}{2}} \cdot \{1 + (p_2 / L)^2\}^{-\frac{1}{2}}.$$

§ 5. EXAMPLES OF RECIPROCAL LATTICE CONSTRUCTION

Two examples are given to illustrate the results and the accuracy obtained in typical cases and to show some features of the line intensities. In these examples, since the camera-length and position of O_1 were not known very precisely, the accuracy obtained in the projections is not the maximum obtainable, but even so, there is no difficulty in recognizing the periodic distribution of the lattice points and the lattice symmetry.

(i) *Rock-salt, NaCl, figure 7(a)*

This relatively simple pattern was obtained from a rock-salt crystal face which had been cut in an arbitrary direction then polished and lightly etched with water. Lines from about fifty planes were visible, some with strong second and higher orders, and as the pattern has a plane of symmetry denoted by the line $m-m$, corresponding planes were numbered similarly, one set with distinguishing dashes. (Since the rock-salt structure is known to have (100) and (110) types of mirror planes the plane "1", i.e. m , must be of this type, and the analysis shows that it was a $(1\bar{1}0)$ plane.) The most prominent zone axis where many of the narrower bands cross, slightly below the centre of figure 7(a), was used for convenience as O_1 , though not a very close approximation to O_1 , and the reciprocal-lattice projection shown in figure 7(b) was constructed from the D and p measurements. The projection, figure 7(b), shows the neighbouring body-centred cubic cells of the reciprocal lattice, and the side elevation on the right demonstrates the rectangularity of the lattice. Probably a better consistency of the coordinates of the points in the outer parts of the projection would have been obtained with the closer estimate of O_1 at about 2 cm. nearer the undeflected-beam spot, and in general it is advisable to use the actual estimated position of O_1 . The Laue indices of the line diffractions relative to the axes chosen are shown in table 2.

The line intensities depend on various factors, including the structure factor, S^2 , which is $16(\text{Na} + \text{Cl})^2$ if h, k, l are all even, $16(\text{Na} - \text{Cl})^2$ when h, k, l are all odd, and zero when h, k, l have other values; h, k, l are the Laue indices and Na and Cl denote the atom-form factors. As a first approximation the atom-form factors can be considered to be represented by the atomic weights multiplied by a function which decreases with increasing deviation of the diffracted beam from the primary beam direction; thus it is natural that the line intensities should decrease with increasing D for both groups, i.e. h, k, l all odd and all even. As an example, in the orders of diffraction from $11\bar{1}$, $11\bar{1}$ has medium, $22\bar{2}$ strong, $33\bar{3}$ nearly zero, and $44\bar{4}$ medium intensity.

The asymmetrical position of the mirror plane m relative to the primary beam enables the relationship between the line intensities and the inclination of the reflecting planes to the beam to be demonstrated clearly in a single reflection pattern by many pairs of band systems. It is possible in this way to see that the expression for the line intensity must contain a factor which reduces the intensity rather rapidly as the angle between the plane and the primary beam (roughly proportional to the perpendicular from the undeflected-beam spot to the band median) approaches zero. Thus, for example, the difference in intensity of the lines of 8 and 8', 7 and 7', 10 and 10' etc., is not large, while in 3 and 3', and 4 and 4' it is larger, and in 5 and 5', 9 and 9', and 11 and 11' it is quite high because the planes 5, 9 and 11' are so much nearer to the primary beam than 5', 9' and 11

respectively. Although the more strongly developed narrow bands almost disappear near the undeflected-beam spot, as in the case of 5, the intensified lines (white in figure 7(a)) of wider bands are often strengthened near the foot of the perpendicular from the undeflected-beam spot as in 24', 25' and 26, apparently because of the narrowness of the selective-reflection range (corresponding to the line sharpness) in such cases, and the steep gradient of the background intensity near the undeflected-beam spot.

Table 2. NaCl line indices and intensities in figure 7(a)

Band no.	Laue indices hkl	Intensity	Band no.	Laue indices hkl	Intensity
1	220	VS	—	—	—
	440	M			
	660	F			
2	111	M	—	—	—
	222	S			
	333	~nil			
	444	M			
3	224	MF	3'	224	MS
4	042	MS	4'	402	VS
	084	VF		804	F
5	131	MF	5'	311	M
	262	M		622	M
6	442	S	—	—	—
7	062	MS	7'	602	S
8	422	S	8'	242	S
	844	VVF		484	VVF
9	282	VF	9'	822	MF
10	420	S	10'	240	MS
	840	F		480	VVF
11	640	M	11'	460	VF
12	462	M	12'	642	M
13	242	S	13'	422	M
14	824	F	14'	284	F
15	464	MF	15'	644	MF
16	311	MS	16'	131	M
	622	MS	—	262	M
17	620	M	—	—	—
18	331	MF			
	662	F			
19	513	F	19'	153	F
20	531	F	—	—	—
21	802	MF	21'	082	—
22	511	F	—	—	—
23	842	VF	23'	482	—
24	064	F	24'	604	M
25	244	F	25'	424	M
26	646	F	26'	466	F
27	664	F	—	—	—

S=strong, M=medium; F=faint; V=very.

A further prominent feature of most Kikuchi-line patterns, conspicuous in figure 7(a), is the peculiar weakening or step-down in intensity of certain sections

of some lines; for example both the white and black lines of 3' almost disappear between the second-order band of 1 and the adjoining first-order band of 4 (or second-order band of 5), while the line-pair of the symmetrical plane 3 behaves similarly across 1 and 4' (not 4). Such conditions help to make the parabolic line envelopes more prominent, for example that (concave to the left) formed by the white lines of 3', 4' (first appearing order), 5' (first-order), 1 (first-order), and the corresponding black lines of 5, 4 and 3, and the mirror-symmetrical parabola (concave to the right) formed by the black lines of the first group and the white lines of the second.

(ii) *Sulphur, figure 8(a)*

This pattern was from an irregular conchoidal fracture of a sulphur crystal of about 1 cm. diameter. The bands are nearly all narrow and correspondingly diffuse-edged, but by applying the criteria described above for the Bragg reflection positions it was possible to measure the band widths to an accuracy of 0.1 mm. and hence to construct quite satisfactorily the projection of the reciprocal lattice, figure 8(b). Here again it is easily seen that the pattern has a plane of symmetry m represented by the median of band 1, with respect to the band positions and intensities; but band 2 represents a plane which is at right angles to 1 but is not a mirror plane.

The lattice-point projections in figure 8(b) lie close to a series of equally-spaced lines parallel to AB. They will not all lie exactly on these lines because the crystal is such that the lattice planes of $(0k0)$ type are not quite normal to the plane of projection, i.e. the photographic plate; in this case, it is sufficiently close for indexing purposes to use these lines without adding all the closely neighbouring lines which would strictly be required in the projection. The points are also grouped on or near lines parallel to CD, which represent projections of lattice rows inclined downward from left to right at about $8\frac{1}{2}^\circ$, as the z coordinates of the points show. The side elevation represents the points of intersection of all these lattice rows with a plane through O at right angles to the projection plane (represented in section by EF) and to CD, and shows the centred rectangle distribution decisively. Strictly the projection of the lattice rows should be done upon a plane which is normal to the rows, not to their projections, but the small inclination of the plane of the illustrated side elevation to this plane introduces little error. The data in figure 8(b) thus show that the Kikuchi-line median positions and the line-pair separations correspond to a face-centred rhombic unit cell, and the ratios $a:b:c$ estimated by calculations based on the measurements of the projection combined with the z coordinates were 0.82:1:1.87. Using the approximately known λL value of 1.13 A. cm. a , b and c were found to be 10.6, 12.9 and 24.2 A. respectively. This is in good agreement with the previous x-ray results, since Mark and Wigner (1924) found rhombic sulphur to have a face-centred rhombic cell with $a=10.61$ A., $b=12.87$ A., $c=24.56$ A. thus $a:b:c=0.820:1:1.900$, with space group V_h^{24} (Fddd) and 128 atoms per cell; Warren and Burwell (1935) who confirmed this space group and calculated the atom positions, found $a=10.48$ A., $b=12.92$ A., $c=24.55$ A., i.e. $a:b:c=0.811:1:1.900$. Trillat and Oketani (1938) concluded from their electron-diffraction transmission patterns that $a=5.2$ A., $b=6.4$ A., $c=12.2$ A., but these values do not take account of spots which occur in some regions of the patterns and show the true Laue-zone spacings, which correspond to cell dimensions close to those of Mark and Wigner (1924) and Warren and Burwell (1935).

Table 3 shows the Laue indices of the observed lines in figure 8(a) and gives, where available, the x-ray structure factor and spot intensity in the rotation pattern, as listed by Warren and Burwell. Unfortunately only eight diffractions

Table 3. Sulphur line indices and intensities in figure 8(a) and x-ray data

Band no.	Laue indices <i>hkl</i>	Int.	X-ray data		Band no.	Laue indices <i>hkl</i>	Int.	X-ray data	
			Struct. factor	Int.				Struct. factor	Int.
1	040	VS	472	S	29	4,4,12	M		
	080	MF	3	nil		(113)	MF		
	0,12,0	VF			30	4,4,12	MF		
	0,16,0	VVF				(113)	F		
2	206	VS	440	VS	31	4,4,10	M		
	4,0,12	F				8,8,20	F		
3	408	M	427	MS	32	4,4,10	VF		
4	131	M			33	4,2,10	M		
	262	MS			34	4,2,10	MF		
	393	M			35	357	VS		
	4,12,4	O			36	357	S		
5	175	S			37	6,2,14	MS		
6	175	S	403	MS	38	6,2,14	S		
7	0,10,2	MF			39	488	MS		
8	0,10,2	F			40	488	MS		
9	062	S	610	S	41	395	M		
	0,12,4	F			42	7,5,15	MF		
10	062	MS			43	5,3,11	M		
	—	—			44	5,3,11	F		
11	151	M			45	6,0,14	M		
12	151	MF			46	3,11,9	F		
13	171	F			47	3,11,9	F		
14	171	F	59	VF	48	6,6,10	M		
15	284	MF			49	3,11,7	F		
16	284	MF			50	5,5,15	F		
17	268	MF			51	135	MS		
18	268	F			52	224	M		
19	4,2,14	F			53	224	F	122	F
20	4,2,14	VF			54	337	MF		
21	359	MS			55	195	F		
22	359	MF			56	195	F		
23	133	F			57	355	M		
24	133	F			58	3,1,11	F		
25	4,2,12	F			59	466	MF		
26	4,2,12	F			60	428	M		
27	319	MS			61	428	F		
28	319	MS			62	193	F		
					63	193	F		

are common to both lists, but these are all in good agreement except in the intensities of successive orders from the (010) plane, where (080) is evidently much stronger in the electron-diffraction pattern than in the x-ray pattern, presumably due to appreciable secondary or dynamic diffraction. Figure 8(a) also shows further examples of the great reduction in intensity of bands which are parallel to the primary beam or nearly so.

§ 6. SUMMARY

The interpretation of electron-diffraction Kikuchi-line patterns from single crystals is developed as a powerful, independent and general means of determining the crystal lattice type, dimensions and orientation and the Laue symmetry of the atomic arrangement. This aim is realized by a construction of the reciprocal lattice by methods which are described and illustrated. The satisfactory accuracy of the method is due to the relatively precise definitions for the location of the Bragg reflection positions, derived and stated here for the first time, which enable the method to be used even when the crystal lattice constants are so large that the bands are narrow and diffuse-edged, as with rhombic sulphur. The patterns shown also illustrate features of the line intensities, especially the reduction in intensity when bands pass near to the undeflected-beam spot.

REFERENCES

- BOERSCH, H., 1937, *Phys. Z.*, **38**, 1000.
 BRANDENBERGER, E., 1937, *Z. Kristallogr.*, **97**, 476.
 EMSLIE, A. G., 1934, *Phys. Rev.*, **45**, 43.
 FINCH, G. I., QUARRELL, A. J., and WILMAN, H., 1935, *Trans. Faraday Soc.*, **31**, 1051.
 FINCH, G. I., and WILMAN, H., 1936, *Proc. Roy. Soc. A*, **155**, 345.
 FINCH, G. I., and WILMAN, H., 1937 a, *Trans. Faraday Soc.*, **33**, 1435.
 FINCH, G. I., and WILMAN, H., 1937 b, *Ergebn. exakt. Naturw.*, **16**, 353.
 HAYASI, T., 1934, *Sci. Rep. Tôhoku Imp. Univ.*, **23**, 491.
 KIKUCHI, S., 1928, *Jap. J. Phys.*, **5**, 83; see also Nishikawa and Kikuchi.
 KIRCHNER, F., 1932, *Ergebn. exakt. Naturw.*, **11**, 64.
 LAMLA, E., 1938, *Ann. Phys., Lpz.*, **32**, 178.
 VON LAUE, M., 1935, *Ann. Phys., Lpz.*, **23**, 705.
 VON LAUE, M., 1936 a, *Ann. Phys., Lpz.*, **25**, 569.
 VON LAUE, M., 1936 b, *Phys. Z.*, **37**, 544.
 MARK, H., and WIGNER, 1924, *Z. phys. Chem.*, **111**, 398.
 VON MEIBOM, R., and RUPP, E., 1933, *Z. Phys.*, **82**, 690.
 NISHIKAWA, S., and KIKUCHI, S., 1928, *Proc. Imp. Acad. Tokyo*, **4**, 475.
 RUTHERFORD, E., and ANDRADE, E. N. DA C., 1914, *Phil. Mag.*, **28**, 263.
 SHINOHARA, K., 1932 a, b, *Sci. Pap. Inst. Phys. Chem. Res. Tokyo*, **18**, 223, 315; 1932 c, *Ibid.*, **20**, 39.
 THIRSK, H. R., and WHITMORE, E. J., 1940, *Trans. Faraday Soc.*, **36**, 565.
 THOMSON, G. P., and COCHRANE, W., 1939, *Theory and Practice of Electron Diffraction* (London: Macmillan).
 TILLMAN, J. R., 1935, *Phil. Mag.*, **14**, 485.
 TRILLAT, J. J., and OKETANI, S., 1938, *Z. Kristallogr.*, **98**, 334.
 WARREN, B. E., and BURWELL, J. T., 1935, *J. Chem. Phys.*, **3**, 6.

Thermionic Emission Constants and Band Overlap

BY E. P. WOHLFARTH

Physics Department, University of Leeds

Communicated by E. C. Stoner; MS. received 15 July 1947

ABSTRACT. A general treatment is given of the effect of band overlap on thermionic emission. The results are applied to nickel, for which the experimental value of the current constant A of $1380 \text{ amp. cm}^{-2} \text{ deg}^{-2}$ differs markedly from the standard value of $120 \text{ amp. cm}^{-2} \text{ deg}^{-2}$. The case of nickel has recently been discussed by Sun Nien T'ai and Band (1946), who derive an expression for the variation of A with temperature. It is shown

that their relation strictly applies only to "holes" in an isolated d -band and is reliable only over a temperature range such that kT/ϵ_{od} is considerably less than unity, ϵ_{od} being the unoccupied width of the d -band. With an overlapping s -band, as in nickel, there is an increasing transfer of electrons from the d -band as the temperature rises. The effect is strongly dependent on the ratio $r_0 - \epsilon_{os}/\epsilon_{od}$, where ϵ_{os} is the occupied width of the s -band. Expressions are derived and precise calculations carried out for the transfer effect and the variation of A over a range $0.0 \leq kT/\epsilon_{od} \leq 1.0$ (including temperatures up to about 2000° K.), and for r_0 values of 10, 20 and ∞ . The last corresponds in effect to an isolated d -band. The results of the calculations are shown in figures 2 and 3. A comparison with the experimental value of A is made using a value of ϵ_{od} estimated from thermal and magnetic data. The satisfactory agreement previously obtained is due to the adoption of a value of the bandwidth which is incompatible with the experimental evidence, and it is shown that the calculated and observed values of A actually differ by a factor of about 6. Various reasons for this discrepancy are discussed, especially the effect of surface impurities.

§1. INTRODUCTION

IN a recent paper Sun Nien T'ai and Band (1946) attempted to interpret anomalous values of thermionic emission current constants on the basis of the electron energy band theory of metals. They consider the effect of band overlap in the vicinity of the surface of the Fermi distribution, and show that under certain circumstances this may lead to abnormally high values of the current constant. In the actual treatment, however, no account seems to be taken of the possibility of a transfer of electrons between the two overlapping bands. The relations obtained are simply those for separate single bands, in each of which there is a fixed number of electrons (or "holes") determined by the degree of overlap. The calculation differs essentially from that for a standard single band only in that the energy density of states is much greater and the width of the unoccupied part of the band much smaller than for a "free-electron" band containing a corresponding number of electrons. To the approximation used, the expressions given appear to be correct, but in substituting a numerical value for the d -band energy width for nickel, the value used differs by a factor of about 10 from that which the most direct interpretation of the available experimental evidence seems to indicate. The agreement between the observed value of the current constant A for nickel (1380 amp. cm⁻² deg⁻² as compared with the standard theoretical value of 120 amp. cm⁻² deg⁻²) and that calculated would thus appear to be largely fortuitous. For this reason, and also because the ideas presented have a bearing on many other physical properties of metals, the problem of thermionic emission constants and band overlap is reconsidered in this paper, with special reference to nickel.

Variations in the value of A , the current constant occurring in the Richardson equation, may be due to a temperature dependence of the thermionic work function χ (Fowler and Guggenheim 1939, p. 484). The Richardson equation, written in the form

$$I = A_0(1 - \bar{r})T^2 \exp(-\chi/kT), \quad \dots\dots(1.1)$$

gives the variation of the emission current per unit area with temperature, \bar{r} being the reflection coefficient, which will be very small for most clean metal surfaces, and A_0 being a universal constant having a value of 120 amp. cm⁻² deg⁻². Hence if χ_0 is the value of χ at absolute zero, and A the value of the current constant at a temperature T , then

$$I = A(1 - \bar{r})T^2 \exp(-\chi_0/kT), \quad \dots\dots(1.2)$$

where

$$A/A_0 = \exp(\chi_0 - \chi)/kT. \quad \dots\dots(1.3)$$

Usually the experimental results are expressed in the form (1.2), with χ_0 constant, so that a decrease of the work function with temperature will be manifested by an increase of A with T . Such a variation will occur for the band form proposed by Mott (1935) and Slater (1936) for nickel. In this assembly the electrons are distributed in two overlapping energy bands, one of which, the “ d -band”, corresponding roughly to the $3d$ electrons, is very much higher than the other one, the “ s -band”, which corresponds to the $4s$ electrons, and which will approximate to a “free-electron” distribution (see figure 1). Fermi-Dirac statistics are readily applicable to this assembly if for the relevant portions of each band $\nu(\epsilon)$, the number of distinguishable states per unit energy range, is proportional to the square root of the energy ϵ . The diagram given by Slater on the basis of numerical calculations indicates that this assumption is justifiable for nickel as giving a reasonable approximation (see § 4). For other metals there may be considerably greater deviation from the standard parabolic band form.

For nickel the equilibrium state at absolute zero is such that the number of electrons in the s -band is equal to the number of absent electrons, or “holes”, in the d -band. Evidence drawn mainly from a study of the magnetic properties of nickel and its alloys indicates that at absolute zero approximately 0.6 electrons per Ni atom are in the s -band, and 9.4 electrons (i.e. $10 - 9.4 = 0.6$ holes) in the d -band, as shown in figure 1. It is through the presence of these holes in the d -band that most of the thermal and magnetic properties of nickel have received satisfactory interpretation.

ϵ_{0s} and ϵ_{0d} are the respective widths of the relevant portions of s - and d -bands at absolute zero. For nickel

$$\int_{\epsilon_{0s}}^{\epsilon_{0d}} \nu_s(\epsilon) d\epsilon = \int_{\epsilon_{0d}}^{\epsilon_{0d}} \nu_d(\epsilon) d\epsilon \quad (1.4)$$

at all temperatures, and it is seen from the diagram that

$$r_0 = \epsilon_{0s}/\epsilon_{0d} \gg 1. \quad (1.5)$$

As will be shown below the equilibrium appropriate to absolute zero will be displaced as the temperature increases, when there will be a transfer of electrons from the d -band to the s -band. As a result the number of holes in the d -band and the number of electrons in the s -band both increase with a rise in temperature. A method of estimating the extent to which electrons are transferred in this way is given in § 2, and it is shown that the current constant as calculated by considering an isolated narrow d -band has to be modified if the transfer effect is taken into account.

§ 2. THEORY OF THE TRANSFER EFFECT IN NICKEL

In this section is given the quantitative treatment of the electron transfer which takes place in nickel. In the analysis the notation employed is in the main that proposed by Stoner (1939 a), whose paper should also be consulted

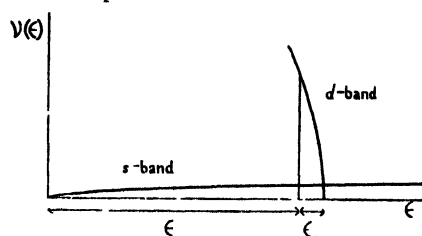


Figure 1. Collective electron assembly for nickel.

ϵ , electronic energy. $\nu(\epsilon)$, energy density of states. ϵ_{0s} , occupied width of the s -band; ϵ_{0d} , unoccupied width of the d -band, both at absolute zero. The diagram is illustrative only (cf. Slater 1936).

for fuller details regarding Fermi-Dirac statistics. The distribution expression applicable to free electrons, or, as here, to the collective electrons in parabolic bands, may be written in the form

$$\frac{dN}{d\epsilon} = \frac{3}{2} \frac{N_0(\epsilon_0^{\frac{1}{2}}/\epsilon_0^{\frac{1}{2}})}{\exp(\epsilon/kT - \eta) + 1}. \quad \dots\dots(2.1)$$

Here N_0 is the number of electrons per atom present in the band and ϵ_0 its occupied width, both at absolute zero. Alternatively N_0 may be the number of holes per atom, in which case ϵ_0 refers to the unoccupied portion of the band. The statistical parameter, η , is related to the chemical potential, or mean Gibbs free energy per particle, μ , by

$$\mu = kT\eta. \quad \dots\dots(2.2)$$

On transforming (2.1), putting $x = \epsilon/kT$, integration over all positive values of x gives

$$N = \frac{3}{2} N_0 \left(\frac{kT}{\epsilon_0} \right)^{\frac{3}{2}} \int_0^{\infty} \frac{x^{\frac{1}{2}} dx}{\exp(x - \eta) + 1}, \quad \dots\dots(2.3)$$

$$\text{or } N = \frac{3}{2} N_0 (kT/\epsilon_0)^{\frac{3}{2}} F(\eta). \quad \dots\dots(2.4)$$

$F(\eta)$ is the Fermi-Dirac integral given in (2.3). The relation (2.4) states essentially that, as the temperature T increases from absolute zero, the total number of particles per atom also changes, unless

$$\frac{3}{2} (kT/\epsilon_0)^{\frac{3}{2}} F(\eta) = 1. \quad \dots\dots(2.5)$$

This relation holds for a single band, not overlapped by any others. In the present case it is satisfied in the limit $r_0 = \epsilon_{0s}/\epsilon_{0d} \rightarrow \infty$. The total number of electrons in the band is then necessarily constant. For the assembly to be considered, however, with r_0 finite, the relation (2.5) does not hold for either band.

For the s -band let N_s and N_{0s} be the number of electrons per atom in the band at temperature T and at absolute zero respectively. Also let ϵ_{0s} be the occupied width of the band (cf. figure 1) and $kT\eta_s$ the chemical potential. Then from (2.4)

$$N_s = \frac{3}{2} N_{0s} (kT/\epsilon_{0s})^{\frac{3}{2}} F(\eta_s). \quad \dots\dots(2.6)$$

A similar expression is applicable for the d -band, but here N_d and N_{0d} are the respective numbers of holes per atom. Let ϵ_{0d} be the unoccupied width and $kT\eta_d$ the chemical potential as measured downwards from the top of the d -band. Denoting the chemical potential of the d -electrons, referred to the same level as that of the s -electrons, namely the bottom of the s -band, by $kT\eta'_d$, one obtains

$$kT\eta'_d = (\epsilon_{0s} + \epsilon_{0d}) - kT\eta_d. \quad \dots\dots(2.7)$$

The relation corresponding to (2.6) is then

$$N_d = \frac{3}{2} N_{0d} (kT/\epsilon_{0d})^{\frac{3}{2}} F(\eta_d). \quad \dots\dots(2.8)$$

For nickel the condition (1.4), which holds at all temperatures, gives

$$N_d = N_s, \quad N_{0d} = N_{0s}. \quad \dots\dots(2.9)$$

The condition of equality of the chemical potential (i.e. of the Gibbs free energy) for thermodynamic equilibrium gives

$$\eta'_d = \eta_s. \quad \dots\dots(2.10)$$

Insertion of the condition (2.7), (2.9) and (2.10) into (2.6) and (2.8) finally yields the relation

$$F\{(\epsilon_{0d} + \epsilon_{0s})/\mathbf{k}T - \eta_d\}/(\epsilon_{0s})^{\frac{1}{2}} = F(\eta_d)/(\epsilon_{0d})^{\frac{1}{2}},$$

or, writing $r_0 = \epsilon_{0s}/\epsilon_{0d}$ and $\tau_0 = \mathbf{k}T/\epsilon_{0d}$,

$$F\{(1+r_0)/\tau_0 - \eta_d\} = r_0^{\frac{1}{2}} F(\eta_d). \quad \dots (2.11)$$

The number of holes in the d -band, N_d , is thus given by (2.8) as

$$N_d/N_{0d} = \frac{2}{3} \tau_0^{\frac{3}{2}} F(\eta_d), \quad \dots (2.12)$$

η_d being determined from (2.11). Owing to the conditions (2.9), the ratio N_s/N_{0s} is equal to N_d/N_{0d} and is thus also given by (2.12). It will be shown below that as T increases the ratio $N_d/N_{0d} = N_s/N_{0s}$ increases from unity at $T=0$. With a rise in temperature the d -band thus loses electrons which are transferred to the s -band. During this process the thermionic work function also changes, and it easy to show from the definition of this quantity (Fowler and Guggenheim 1939, p. 477) that

$$(\chi_0 - \chi) = -(\mathbf{k}T\eta_d - \epsilon_{0d}) = (\mathbf{k}T\eta_s - \epsilon_{0s}). \quad \dots (2.13)$$

Hence from (1.3)

$$A/A_0 = \exp(\tau_0^{-1} - \eta_d). \quad (2.14)$$

Precise numerical calculations of N_d/N_{0d} and of A/A_0 , using (2.12) and (2.14), have been carried out for a series of temperatures (§ 3). These may be usefully supplemented by series expansions appropriate for $\tau_0 \ll 1$. For $\tau_0 \ll 1$, $\eta_d \gg 1$, $\{(1+r_0)/\tau_0 - \eta_d\} \gg 1$, the asymptotic series expansion of the Fermi-Dirac integral may be written in the form

$$F(\eta_d) = \frac{2}{3} \eta_d^{\frac{3}{2}} \{1 + (\pi^2/8)\eta_d^{-2} + \dots\}.$$

Similarly, and to the same degree of approximation, writing

$$z = \{(1+r_0)/\tau_0 - \eta_d\},$$

$$F(z) = \frac{2}{3} z^{\frac{3}{2}} \{1 + (\pi^2/8)z^{-2}\}.$$

Hence from (2.11), after some rearranging,

$$\tau_0^{-1} - \eta_d = \frac{\pi^2}{12} \left[\frac{r_0/\tau_0 - \eta_d}{\eta_d \{(1+r_0)/\tau_0 - \eta_d\}} \right]. \quad \dots (2.15)$$

To a first approximation the right hand side of (2.15) is zero, so that $\eta_d = \tau_0^{-1} = \epsilon_{0d}/\mathbf{k}T$. Inserting this value into the right hand side the second approximation is obtained in the form

$$\tau_0^{-1} - \eta_d = (\pi^2 \tau_0/12)(r_0 - 1)/r_0. \quad \dots (2.16)$$

Hence from (2.8) and (2.9)

$$N_d/N_{0d} = N_s/N_{0s} = 1 + (\pi^2/8r_0)\tau_0^2. \quad \dots (2.17)$$

The relation (2.17) shows that for $\tau_0 > 0$, $N_d > N_{0d}$ and $N_s > N_{0s}$, so that there is in fact a transfer of electrons from one band to the other. The temperature variation of the current constant is given by (2.14) and (2.16) as

$$A/A_0 = \exp\{(\pi^2 \tau_0/12)(r_0 - 1)/r_0\}. \quad \dots (2.18)$$

Since τ_0 will in most cases of physical interest rarely exceed unity (see § 4) it is unnecessary to derive any high temperature approximations, applicable if $\tau_0 \gg 1$. A point of interest does, however, arise in connection with (2.11). For $\tau_0 \rightarrow \infty$ this relation defines a certain lower limit of η_d , which is here denoted by y , where

$$F(-y) = r_0^{\frac{1}{2}} F(y). \quad \dots (2.19)$$

For $r_0 > 1$, y will be negative and will be a function of r_0 only. The expression (2.14) now shows that the limiting value of A/A_0 (for $\tau_0 \rightarrow \infty$) is very simply related to y by

$$A_{\max}/A_0 = \exp|y|, \quad \dots\dots(2.20)$$

A_{\max} being the maximum value of the current constant for any particular value of r_0 . Values of y and A_{\max}/A_0 have been calculated for $r_0 = 10$ and 20 and for $r_0 \rightarrow \infty$ and are given in § 3 as supplementing the curves shown in figure 3.

As regards the relations derived by Sun Nien T'ai and Band (1946) it may be noted that, since $\tau_0 = \hbar T/\epsilon_{0d}$ the low temperature expression (2.18) is exactly equivalent to their relation (13), which, assuming that $r_0 \gg 1$, was further simplified to give

$$A/A_0 = \exp\{(\pi^2/12)(\hbar T/\epsilon_{0d})\}. \quad \dots\dots(2.21)$$

This expression was taken to be the basic relation and was employed for "complete calculations of current-temperature curves". Such a procedure is, however, open to objection. In the first place the low temperature approximations used are reliable only if $\hbar T/\epsilon_{0d} \ll 1$. For nickel, however, ϵ_{0d}/\hbar is of the order of 10^3 °K. (see § 4), so that, at the higher temperatures used in experimental determinations of the thermionic constants, $\hbar T/\epsilon_{0d}$ may be of the order unity. For this reason precise calculations of the type outlined in § 3 have to be carried out. Furthermore, by stating their relation (14) in the form given, Sun Nien T'ai and Band seem to neglect the effect of the s -band altogether. This relation would in fact be obtained by considering the temperature variation of $(\epsilon_0/\hbar T - \eta)$ for the d -band by itself, the transfer effect being thus neglected. A second approximation to the Richardson equation had, in fact, already been derived in this way by Fowler and others (see Reimann 1934), although, since particles rather than holes were considered, a decrease of A with temperature had been predicted.

§ 3. COMPUTATIONAL DETAILS AND RESULTS

The equations to be solved are given in § 2 and are here rewritten for convenience:—

$$F\{(1+r_0)/\tau_0 - \eta_d\} = r_0^{\frac{1}{2}} F(\eta_d), \quad \dots\dots(3.1)$$

$$N_d/N_{0d} = \frac{3}{2} \tau_0^{\frac{1}{2}} F(\eta_d), \quad \dots\dots(3.2)$$

$$A/A_0 = \exp\{\tau_0^{-1} - \eta_d\}. \quad \dots\dots(3.3)$$

As has been stated the approximations (2.17) and (2.18) are applicable only if $\tau_0 \ll 1$. Since, however, this condition is not satisfied if the temperature is high and if ϵ_{0d} is relatively small, it may be necessary to obtain exact numerical solutions of the equations for a suitable range of τ_0 , say $0.0 \leq \tau_0 < 1.0$, and for one or more values of r_0 . For nickel the magnitude of this parameter is rather uncertain. This limitation of the present treatment is discussed further in § 4, where estimates of ϵ_{0s} and ϵ_{0d} are obtained. These indicate that r_0 lies between 10 and 40. The calculations in this section were carried out for two values of r_0 , namely 10 and 20. Additional calculations for $r_0 \rightarrow \infty$, corresponding to an infinitely wide s -band, can be made very simply, and the diagrams showing the results of the three sets of calculations almost certainly cover the physically relevant range.

It would be most convenient to obtain values of N_d/N_{0d} and of A/A_0 for equally spaced values of τ_0 lying in the desired range. For the present purpose

it was found to be simpler, however, to proceed as follows. For a given value of r_0 a trial value of η_d is first chosen, satisfying the condition that $\eta_d > y$, where y is given by (2.19). The magnitude of the corresponding Fermi-Dirac integral $F(\eta_d)$ is then obtained from the tables of McDougall and Stoner (1938), so that $F\{(1+r_0)/\tau_0 - \eta_d\}$ is calculable from (3.1). To obtain values of the argument inverse interpolation in the same tables would be necessary if r_0 were small. For r_0 equal to 10 or 20, however, the left hand side of (3.1) is large, and the value of $\{(1+r_0)/\tau_0 - \eta_d\}$ can be obtained from the inverse asymptotic series, which is discussed fully by McDougall and Stoner. With

$$\alpha = [\frac{3}{2}F\{(1+r_0)/\tau_0 - \eta_d\}]^{-\frac{1}{3}} = r_0^{-1} [\frac{3}{2}F(\eta_d)]^{-\frac{1}{3}}$$

it is shown that

$$\{(1+r_0)/\tau_0 - \eta_d\} = \alpha^{-1} \{1 - a_2\alpha^2 - a_4\alpha^4 - a_6\alpha^6 - \dots\}, \quad \dots\dots(3.4)$$

where $a_2 = 0.822\,467$, $a_4 = 1.217\,614$, $a_6 = 9.161\,386$.

For the purpose of the present calculations the three terms of the series given in (3.4) were found to give a sufficiently high accuracy. Having obtained $\{(1+r_0)/\tau_0 - \eta_d\}$ from (3.4), η_d is calculated using the same value of r_0 , 10 or 20, as before. If this value of τ_0 lies outside the range to be covered ($0.0 < \tau_0 \leq 10$), a different value of η_d has to be chosen and the calculation repeated. With a suitable η_d , τ_0 having been obtained by the method outlined, N_d/N_{0d} and A/A_0 are determined from (3.2) and (3.3). About 10 values of τ_0 lying in the desired range were calculated for $r_0 = 10$ and 20 respectively, the computations being carried out with the aid of a Brunsviga calculating machine, using Barlow's Tables (ed. Comrie 1941) for the roots and the Federal Works Agency Tables (1939) for the exponentials. The procedure may be illustrated by giving the details for $r_0 = 10$, $\eta_d = 1.0$, the numbers being given to five significant figures. Then

$$\begin{aligned} F(\eta_d) &= 1.3964, & \tau_0 &= 0.63510, \\ F\{(1+r_0)/\tau_0 - \eta_d\} &= 44.157, & N_d/N_{0d} &= 1.0601, \\ \{(1+r_0)/\tau_0 - \eta_d\} &= 16.320, & A/A_0 &= 1.8766. \end{aligned}$$

The final results are most conveniently given in graphical form and are shown in figures 2 and 3. Besides the results for $r_0 = 10$ and 20, those for $r_0 \rightarrow \infty$ are also indicated. These were calculated using the relation (2.5) which gives

$$\left. \begin{aligned} \tau_0 &= \{\frac{3}{2}F(\eta_d)\}^{\frac{1}{3}}, \\ N_d/N_{0d} &= 1, \\ A/A_0 &= \exp[\{\frac{3}{2}F(\eta_d)\}^{\frac{1}{3}} - \eta_d] \end{aligned} \right\}. \quad \dots\dots(3.5)$$

Figure 2 shows the variation of N_d/N_{0d} with temperature. As is to be expected from the relations (2.5) and (3.2), the transfer effect will decrease in magnitude as r_0 increases, until, when $r_0 \rightarrow \infty$, the relation (2.5) is obeyed exactly so that N_d/N_{0d} is unity at all temperatures. The physical explanation is that, since the number of electrons per atom in the s -band at absolute zero remains constant at 0.6, an increase of ϵ_0 , relative to ϵ_{0d} implies a corresponding decrease in the height $\nu_s(\epsilon_0)$, so that it becomes increasingly more difficult for a d -electron to enter the s -band. For the smaller values of r_0 shown in the diagram the electron transfer may become quite considerable. For $r_0 = 10$ and $\tau_0 = 1.0$, to consider an extreme example N_d/N_{0d} is as much as 1.14, so that the number of holes per atom at this temperature will have increased from 0.6 to nearly 0.7.

The possible applications of this effect to other physical properties have been discussed elsewhere.

In figure 3 the three curves give the variation of A/A_0 with temperature for the values of τ_0 indicated. It will be noticed that A/A_0 is much less dependent on the value of τ_0 than is N_d/N_{0d} , and, furthermore, that for any given value of τ_0 , A/A_0 increases with τ_0 , reaching a maximum when $\tau_0 \rightarrow \infty$, i.e. if the d -band is overlapped by an infinitely wide and low s -band. In that case the increase of the current constant is due entirely to a normal decrease, with temperature, of the work function as given by the relations (2.5) and (2.13). For finite values of τ_0 , however, the transfer of electrons from the d -band to the s -band causes a relative increase of the work function owing to the tendency of the Fermi limit

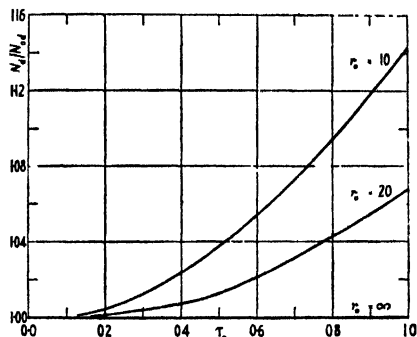


Figure 2. Transfer effect for nickel

reduced temperature kT/ϵ_{0d} . N_d , N_{0d} , number of holes in d -band at reduced temperature τ_0 and at absolute zero respectively. The curve for $\tau_0 = \infty$ coincides with the τ_0 -axis.

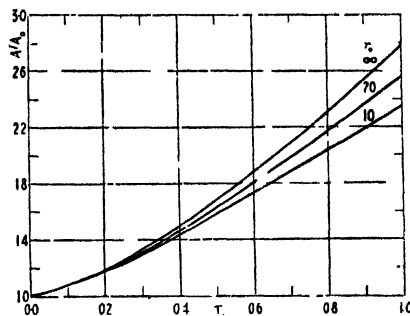


Figure 3. Temperature variation of A/A_0 for nickel.

τ_0 , reduced temperature kT/ϵ_{0d} . A , A_0 , thermionic emission current constant at reduced temperature τ_0 and at absolute zero respectively. $A_0 = 120$ amp. cm^{-2} deg^{-2} .

to shift to regions of lower energy (see figure 1). Hence the current constant will decrease relative to its particular value for $\tau_0 = \infty$. In the extreme case of $\tau_0 = 1$ the increase in the work function due to the transfer effect just balances the normal temperature decrease, and $A/A_0 = 1$ at all temperatures. This is at once apparent from the relation (3.1), which gives, with $\tau_0 = 1$, $\eta_d = \tau_0^{-1}$, so that, from (3.3), $A/A_0 = 1$. As has been stated in § 2 an increase of the current constant will occur only if holes are relevant, while for particles A/A_0 will decrease with temperature. For any finite value of τ_0 , A/A_0 will tend asymptotically to a limiting value A_{max}/A_0 as $\tau_0 \rightarrow \infty$, as is indicated by the discussion given in § 2, where it was shown (equation (2.20)) that $A_{\text{max}}/A_0 = \exp|y|$. Calculations of y , using (2.19) can be readily carried out by employing a modified Taylor's series method, the required derivatives of $F(\eta)$ being given by McDougall and Stoner (1938). It is found that for

$$\tau_0 = 10, \quad y = -2.2527, \quad A_{\text{max}}/A_0 = 9.5134$$

$$\text{and for } \tau_0 = 20, \quad y = -2.9844, \quad A_{\text{max}}/A_0 = 19.775.$$

These figures indicate that A_{max}/A_0 increases with τ_0 , until, for $\tau_0 \rightarrow \infty$, $A_{\text{max}}/A_0 \rightarrow \infty$.

It is of interest in this connection to compare the results obtained by the methods of this section with those obtainable from the low temperature expression (2.18). Selecting a particular value of τ_0 , say $\tau_0 = 1.0$, the three values of A/A_0

determined roughly from the curves of figure 3 are 2.35 ($r_0 = 10$), 2.56 ($r_0 = 20$) and 2.79 ($r_0 \rightarrow \infty$), while the corresponding values as calculated from (2.18) are 2.10, 2.18, and 2.28 respectively. The magnitude of A/A_0 given by (2.21) (equivalent to equation (14) of Sun Nien T'ai and Band's paper) is equal to that just given for $r_0 \rightarrow \infty$, i.e. $A/A_0 = 2.28$. The discrepancy between the two values of A/A_0 as calculated by the exact methods of this section and from the approximate relation is thus not very large, even for $\tau_0 = 1.0$. The considerations given in the preceding section show, however, that use of the approximation is nevertheless not justified, except when $\tau_0 \ll 1$.

§4. COMPARISON WITH EXPERIMENT: CONCLUSION

In order to compare the results deduced from the theoretical considerations of the preceding section with those that have been obtained by experiment, it is necessary to make an estimate of the bandwidth ϵ_{0d} for nickel. Two methods of estimating this quantity from observational data may conveniently be employed. In the first place, if the low temperature specific heat of a metal can be expressed in the form

$$c_v = \gamma T + bT^3,$$

the linear term gives the contribution of the electrons, and it has been shown (Stoner 1939 b), that for parabolic bands in ferromagnetic metals γ may be estimated from the relative magnetization at absolute zero, ζ_0 , and the bandwidth ϵ_{0d} . For nickel a consideration of the magnetic properties indicates that ζ_0 is very nearly equal to unity, so that from the experimental value of γ , namely $17.44 \cdot 10^{-4}$ cal deg⁻² mol⁻¹ (Keesom and Clark 1935), ϵ_{0d}/k is found to be about 2200° K., provided account is taken of the small contribution of the *s*-electrons. Alternatively ϵ_{0d}/k may be estimated from the observed Curie temperature, using the relation (Stoner 1938)

$$\epsilon_{0d}/k = \theta_c f(\zeta_0),$$

applicable to parabolic bands, for which $f(1.0) = 0.41$. Hence with $\theta_c = 631$ K., ϵ_{0d}/k is found to be about 1600° K. The value of ϵ_{0d}/k may be estimated on the assumption that the *s*-electrons are "free", as in copper, giving $\epsilon_{0s}/k = 6 \cdot 10^4$ K., which has to be regarded as an upper limit. It is difficult to choose between the two estimates of ϵ_{0d}/k given above and also to see how far the premises underlying the calculation of ϵ_{0d}/k are correct. The main difficulty appears to be the uncertainty as to deviations of actual band shapes from that postulated, and as to the different effects such deviations may have on the various theoretically determined magnetic and energetic properties of metals. For the electron assembly considered both the *d*-band and the *s*-band may conceivably deviate from a parabolic shape, and the extent of the deviations could only be gauged if the calculations of Slater (1936) were to be carried further with an aim to obtaining a higher accuracy in the final results.

It is fairly certain, however, that the estimate of ϵ_{0d}/k given by Sun Nien T'ai and Band is far too low. In their paper it is stated that "in fact electronic specific heat data give ϵ_{0d} as small as 0.025 e.v.", corresponding to $\epsilon_{0d}/k = 300^\circ$ K. The calculation outlined above, on the other hand, shows that the bandwidth deducible from the low temperature specific heat is 2200° K. Using their value of ϵ_{0d}/k , giving, for $T = 1400^\circ$ K., $\tau_0 = 4.7$, Sun Nien T'ai and Band were able to explain the anomaly in the observed current constant for nickel quite well. However,

not only is the estimate of the bandwidth erroneous, but the low temperature expression for A/A_0 will certainly not be applicable for $\tau_0 \simeq 5$. It is easy to see in fact that the agreement between theory and experiment is much more unsatisfactory. The most recent observed value of A is 1380 amp. cm⁻² deg⁻² (Fox and Bowie 1933), the experimental results used for its calculation being obtained at a mean temperature of about 1400° K. Assuming $\epsilon_{od}/k = 2000^\circ$ K., i.e. $\tau_0 = 0.7$, figure 3 shows that A/A_0 will be of the order of 2.0, i.e. $A \simeq 240$ amp. cm⁻² deg⁻². This discrepancy may be due to a number of causes, a summary of which has been given by Reimann (1934). Among these are:

(i) Surface charging effects. Theoretical discussions of these effects have been given in a number of papers, references to which may be found in Reimann's book. Since, however, the electrons emitted by hot bodies originate in the interior of the emitter, surface effects of this nature may be assumed to be of secondary importance only (see Reimann, p. 271)

(ii) Variation of emitter area. Since any metal surface will in general not be perfectly smooth the actual emitter area will be larger than that calculated geometrically. This would imply an increase of the current constant above the value appropriate to the observed area.

(iii) Thermal expansion effects. These are too complex to discuss fully and reference should be made to the original papers.

(iv) The main reason for the discrepancy appears almost certainly to be the imperfect purity of the emitter surface. As is well known even monatomic impurity layers may influence the thermionic emission very appreciably. This is borne out by the thermionic properties of platinum, which were first investigated by Du Bridge (1928) who found a value of A equal to 17 000. Later work by Van Velzer (1933), who subjected his specimen to some very prolonged heat treatment, showed that for platinum A was very strongly dependent on the presence of gaseous impurities on the surface and in the surrounding atmosphere. The final value of A obtained by Van Velzer was equal to 170, and, it is stated, "slightly lower values of A might have been obtained had the filament withstood further treatment". Fox and Bowie took great care to treat their nickel specimen thoroughly before the investigation. As Van Velzer had shown, however, a very stable state which resisted aging at 1650° K. was attained during the treatment of his platinum specimen. For this state A was abnormally high. Since all the heat treating processes for nickel were carried out at temperatures below 1650° K., the anomalous result for this metal might also be partly indicative of such a state, corresponding to one of imperfect surface purity.

ACKNOWLEDGMENT

The problems discussed in this paper were considered in the course of a more general investigation of the effect of band overlap on the properties of nickel and its alloys with copper, more particularly the thermal and magnetic properties.

The author wishes to express gratitude to Professor E. C. Stoner for his guidance in this and other work.

REFERENCES

- COMRIE, L. J. (ed.), 1941, *Barlow's Tables* (London: Spon).
 DU BRIDGE, L. A., 1928, *Phys. Rev.*, **32**, 961.
 FOWLER, R. H., and GUGGENHEIM, E. A., 1939, *Statistical Thermodynamics* (Cambridge: The University Press).

- FOX, G. W., and BOWIE, R. M., 1933, *Phys. Rev.*, **44**, 345.
 KEESOM, W. H., and CLARK, C. W., 1935, *Physica*, **2**, 513.
 LOWAN, A. N., *et al.*, 1939, *Tables of the Exponential Function* (New York : Works Project Administration).
 McDougall, J., and STONER, E. C., 1938, *Phil. Trans. Roy. Soc. A*, **237**, 67.
 MOTT, N. F., 1935, *Proc. Phys. Soc.*, **47**, 571.
 REIMANN, A. L., 1934, *Thermionic Emission* (London : Chapman and Hall).
 SLATER, J. C., 1936, *Phys. Rev.*, **49**, 537.
 STONER, E. C., 1938, *Proc. Roy. Soc. A*, **165**, 372 ; 1939 a, *Phil. Mag.*, **28**, 257 ; 1939 b, *Proc. Roy. Soc. A*, **169**, 339.
 SUN NIEN T'AI, and BAND, W., 1946, *Proc. Camb. Phil. Soc.*, **42**, 72.
 VAN VELZER, H. L., 1933, *Phys. Rev.*, **44**, 831.

The Barkhausen Effect.

By H. D. BUSH AND R. S. TEBBLE

Department of Physics, The University of Leeds

Communicated by R. Whiddington; MS. received 10 August 1947

ABSTRACT. A new method of investigating the discontinuous changes in magnetization (the Barkhausen effect) is described. The number of discontinuities occurring is counted, and a knowledge of the change in magnetic moment associated with each one, leads to information concerning the volume of material associated with a discontinuity. The change in magnetization due to discontinuities and the distribution of number and size over the hysteresis cycle is obtained. Conclusions are drawn that the discontinuities associated with volumes of greater than $0.9 \cdot 10^{-9} \text{ cm}^3$ only account for about 30% of the total change in magnetization. Possible explanations of these results are discussed, and suggestions made as to the further development of the method.

§1. INTRODUCTION

It is well known that on the steep part of the hysteresis curve for ferromagnetics, the change in magnetization proceeds discontinuously (the Barkhausen effect) and that the discontinuous changes in flux can be amplified as voltage pulses, and displayed on an oscilloscope or made to operate a loud-speaker, when the familiar "rustling" sound can be heard.

The discontinuities have been identified with changes in the directions of magnetization of the domains of ferromagnetic theory. The volume of material constituting a domain and the extent to which the discontinuous processes account for the change in magnetization, are matters of considerable interest, and have been investigated by various observers.

Barkhausen (1919) in his original paper stated that the larger discontinuities corresponded to the reversal of the magnetic moment in volumes which can be estimated to be between 10^{-5} and 10^{-6} cm^3 . Tyndall (1924) also found that the larger discontinuities corresponded to a volume of 10^{-6} cm^3 , and that the period of the discontinuity was about 10^{-3} seconds.

Bozorth and Dillinger (1929, 1930, 1932) in a series of comprehensive experiments, examined the size of discontinuities in wire specimens, together with their distribution over the hysteresis cycle. They came to the following conclusions: (a) The average volume of discontinuities varies over different parts of the field

range, the maximum occurring in the steep part of the hysteresis curve. (b) For annealed iron, hard-drawn iron, hard-drawn permalloy and annealed nickel, the discontinuities corresponding to volumes greater than 10^{14} atoms (10^{-9} cm³), accounted for most of the change in magnetization over the steep part of the hysteresis curve. (c) The maximum average volume ranged from 1.2×10^{-9} cm³ for annealed iron to 4.5×10^{-9} cm³ for malloy. (d) It was suggested that these volumes could be identified with the ferromagnetic domains of magnetization.

Bozorth and Dillinger's method was to amplify the voltage pulses produced in a search coil by the discontinuities. From the output current of the amplifier, dB/dt , the rate of change in induction in the specimen caused by discontinuities could be calculated.

Knowing dH/dt , the rate of change of magnetizing field, dB/dH could be found. dB/dH , given by this method of integrating the separate pulses, was compared with the value of dB/dH found from the hysteresis curve.

The average volume of material \bar{v} , the magnetization of which changes from saturation in one direction to saturation in the other, is obtained from an expression giving \bar{v}^2/\bar{v} in terms involving, among other factors, the mean square current of the amplifier, the reversible permeability and the conductivity of the specimen.

With the development of modern methods of counting electrical pulses, it was thought that it would be possible to count the number of discontinuities of various sizes which occur when a ferromagnetic is slowly magnetized, so that distribution curves of number against size could be obtained. Further curves could be derived showing the contribution to the magnetization of discontinuities of various sizes, and also the total change in magnetization caused by the total discontinuities detected. This appeared to be a very direct method of approach to the problem.

In this preliminary paper, a description is given of the apparatus developed for this purpose, together with some typical results. The work shows that the method is entirely practicable, and interesting results have already been obtained. The character of the results indicates the directions in which more detailed measurements are required and the developments of the method which are desirable. The results of more comprehensive investigations will, it is hoped, be presented in due course.

§ 2. APPARATUS

A block diagram of the apparatus is shown in figure 1. The two search coils are connected in series opposition and the specimen is placed inside one, this assembly being enclosed in the magnetizing coil. By altering the current in the magnetizing coil by means of a current changing device, the magnetizing field can be slowly changed at an approximately uniform rate.

The changes in flux caused by Barkhausen discontinuities are picked up by the appropriate search coil, and amplified as voltage pulses by the amplifier. The discriminator selects pulses of amplitude greater than any desired value, and these are then counted by a scale-of-sixteen unit and a mechanical recording counter.

The characteristics of the apparatus are as follows. The magnetic field changing unit enables the current in the magnetizing coil to be slowly changed. This is done by charging and discharging a condenser, arranged to alter the grid bias of a pentode valve, thus causing the anode current to change at the required rate. The unit is based on a circuit described by Bozorth (1929). The field can be changed at a rate of 2 to 20 milli-oersteds per second for a maximum field of

20 oersted. For smaller maximum fields the rate of change is proportionately less; in the case of hard-drawn mumetal it was 0.15 to 1.5 milli-oersted per second for a maximum field of 2.6 oersted.

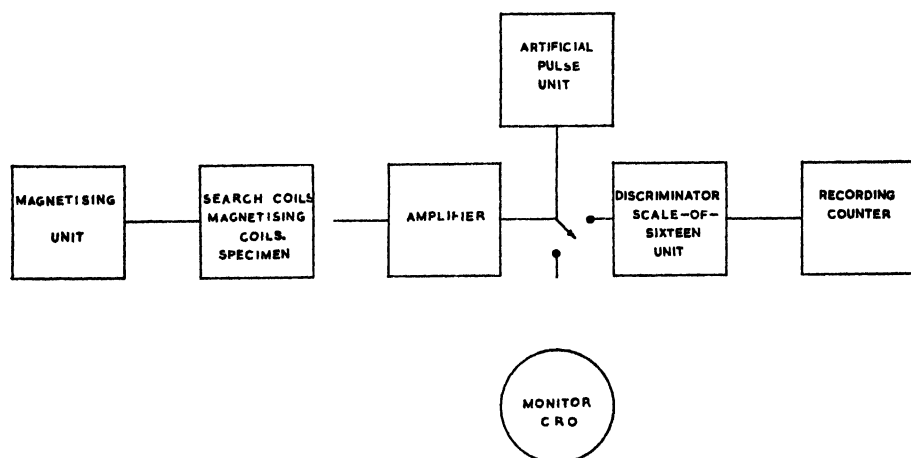


Figure 1. Block diagram of the apparatus

The amplifier is of conventional design, consisting of five stages of resistance capacity coupling. The noise level with the search coils connected was about $2.5 \mu\text{V}$, the first valve being carefully chosen to generate as little noise as possible. The frequency response is shown in figure 2.

The discriminator is based on a design due to Lewis (1942) and selects pulses above any desired size, enabling the various magnitudes of change in magnetic moment, corresponding to Barkhausen discontinuities, to be measured. The discriminator would differentiate between pulses whose amplitudes differed by only 0.5 volts. The scale-of-sixteen unit is built up from the hard valve scale of two circuits originally described by Stevenson and Getting (1937). This scaling unit enabled the actual counting to be done by a recording counter of the watch type (Neher 1939), which had a minimum resolving time of about $1/300$ second. A special driving unit was used to operate the recording counter.

An artificial pulse unit was constructed which gave square pulses of about the same time duration as those produced by Barkhausen discontinuities; this was used to calibrate the amplifier and to test the counting circuits. It is of conventional design; the pulse recurrence frequency is determined by a resistance capacity phase shift oscillator, the sine-wave output being fed into a squaring valve. The square wave so produced is passed into a circuit of short time constant for differentiation. The negative and positive pulses produced are passed into a third valve which removes the positive pulse by grid current action, and amplifies and squares the negative one to give a positive pulse output. Variation of the short time constant for differentiation provides a control of pulse duration. The low impedance output is provided by a cathode follower. The

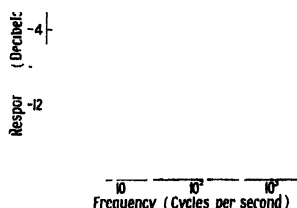


Figure 2. Frequency response of amplifier.

pulse recurrence frequency is variable from 100 to 200 cycles per second and the pulse duration has a range of 300 to 1000 microseconds.

The apparatus can detect changes in magnetic moment of greater than 3×10^{-8} E.M.U. Thus if these changes correspond to complete reversals of the direction of magnetization of domains, the apparatus will detect domain volumes greater than 0.9×10^{-9} cm³ in the case of iron where the saturation intensity I_0 is about 1.7×10^3 . The domain volume is given by the equation:

$$2I_0 = \Delta M/V,$$

where ΔM is the change in magnetic moment in a volume V . Pulses can be counted at a maximum speed of 5000 per second.

The essential limitation to the sensitivity of the apparatus is set by the noise generated in the amplifier, a large part of which is caused by thermal agitation noise set up by the resistance of the search coils. Hence in the design of the search coils it is important to keep the resistance as low as possible, consistent with the necessity of obtaining the maximum sensitivity. Each search coil consists of 50 000 turns of 44 s.w.g. copper wire, wound in the form of a multi-layer solenoid, 10 cm. long, 0.68 cm. internal diameter and 1.63 cm. external diameter. With these dimensions

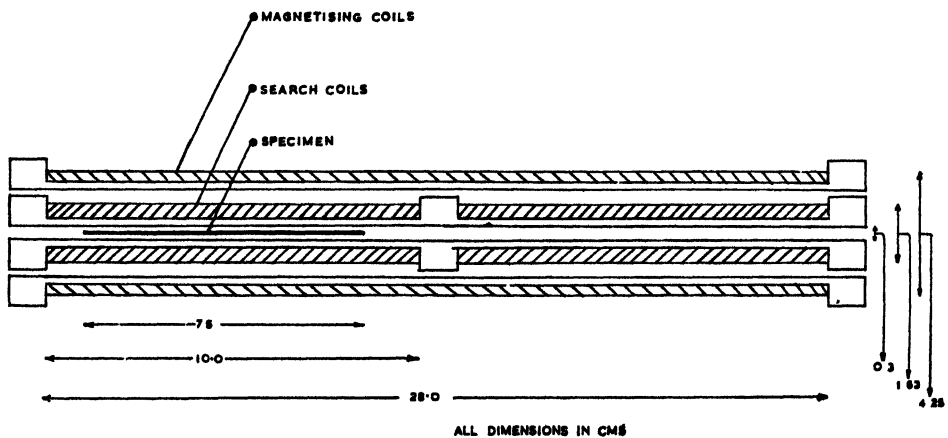


Figure 3. Layout of specimen, search coils, and magnetizing coil

the response to the changes in flux over most of the length is sensibly constant. Two search coils in series opposition are used to reduce the effect of external interference. Figure 3 shows a section through the assembly of search coils, magnetizing coil, and specimen. The assembly is enclosed in a mumetal box which also reduces the effects of external interference and the effect of stray magnetic fields. With the specimen in the form of a wire 7.5 cm. long placed in the position indicated in figure 3 only one search coil is affected by Barkhausen discontinuities taking place in the specimen.

A fuller description of the apparatus is given in a thesis by one of us (H.D.B. 1947).

§ 3. CALIBRATION OF THE APPARATUS

The time duration of a Barkhausen reversal, as seen in the output of the amplifier, is of the order of 300 microseconds, and depends on the time constant of the search coil and amplifier, and on the rate of decay of eddy currents in the

specimen. However, the actual change in magnetic moment, and resultant initial change in flux, take place in a time which is short compared with the time constant of the apparatus. Under these last conditions, the amplitude of the resultant output pulse, in the case of a linear amplifier, should be proportional to the change in flux in the search coil.

A calibrating coil, 5 cm. diameter, 10 cm. long and two turns per cm., was placed over the search coil which normally contained the specimen. The flux change produced in the search coil by a known current passing through the calibrating coil could be calculated, making allowance for the variation in field along the length of the coil and over the cross section of the search coil. The relation between flux change in the search coil, and the change in magnetic moment resolved in a direction parallel to the applied field, for a discontinuity taking place in the specimen, is given by:

$$\text{Flux change} = 2\pi\Delta MN / (r^2 + l^2)^{\frac{1}{2}},$$

where ΔM = change in magnetic moment;

N = number of turns in search coil;

r = effective radius of search coil;

$2l$ = length of search coil.

This formula applies strictly only for a discontinuity taking place at the centre of the search coil. However, as the ratio of length to diameter of the search coil is 9:1, the mean variation in sensitivity over the length occupied by the specimen (3/4 that of the search coil) is only 1%.

The change in flux was produced by passing a current pulse of square wave form (from the artificial pulse unit), through the calibrating coil, with various high resistances in series. In this way it was possible to detect the variation in gain of the amplifier caused by non-linearity, and any effect of leakage across the series resistance.

The characteristics of the search coil were such that the rapid change of flux tended to set up an oscillation in the search coil, at a frequency determined by its self-inductance and the associated circuit capacities. There was sufficient damping across the coil to prevent more than one cycle of oscillation. Experiment showed that the amplitude of the voltage pulse so produced was proportional to the change in flux, but independent of its duration over a time range of three to one (300 to 1000 $\mu\text{sec.}$). Pulses from the amplifier were fed into the discriminator, and the setting of the discriminator required to cut off pulses of a given size determined. In this way it was possible to obtain a calibration curve showing variation of change in magnetic moment with the corresponding discriminator setting.

The counting system was tested for accuracy and speed of counting by means of the pulse unit; with sinusoidal pulses, counting speeds up to 5000 cycles per second are possible.

§ 4. EXPERIMENTAL METHOD

The discriminator was set at a given level; the specimen was then magnetized to saturation in one direction and the field gradually reduced, reversed and then increased to saturation in the opposite direction. Readings of the recording counter were taken, in the case of iron at intervals of 2 oersteds, and in the case of other specimens at other suitable intervals. This was repeated a number of times

at each level of the discriminator and an average taken in each case. A curve was thus obtained for each specimen giving the number of discontinuities N above a given minimum size ΔM plotted against size, as in figure 8 for annealed iron. By differentiating this curve it was possible to obtain a size/number (ΔM vs $dN/d\Delta M$) distribution curve and hence a curve giving size, ΔM , plotted against contribution, $\Delta M \cdot dN/d\Delta M$; see figures 4, 6, 8 and 5, 7, 9 respectively.

The area under the contribution curve

$$\int \Delta M \cdot \frac{dN}{d(\Delta M)} \cdot d(\Delta M),$$

with appropriate limits, is the total change in magnetic moment contributed by discontinuities which can be detected above the noise level of the amplifier. This total divided by the volume of the specimen gives the change in intensity, contributed by the Barkhausen discontinuities measured, over the part of the hysteresis cycle considered.

When the discriminator is set near the noise level there is a background count, constant for each setting, which is appreciable and increases as the noise level is approached. If this becomes comparable with the true count, the error in counting becomes large. It was found that accurate results could not be obtained with a background count of greater than 10 per second at the recording counter. A lower limit to the discriminator setting was therefore put at $\Delta M = 3.0 \times 10^{-6}$ E.M.U.; above the setting $\Delta M = 3.75 \times 10^{-6}$ E.M.U. the background count was negligible.

§ 5. RESULTS

The following materials, in the form of thin wires 7.5 cm. long, were examined:—

	Diam. cm.	Composition
Annealed iron	0.0376	99.92% Fe, 0.03% C, 0.05% Mn.
Hard-drawn iron	0.0376	
Hard-drawn nickel	0.0458	99.5% Ni.
Hard-drawn mumetal	0.0254	76% Ni, 17% Fe, 5% Cu, 1.5% Cr., 0.6% Mn.

The hysteresis curves for these materials are shown in figures 4 to 7. These curves were obtained using the standard ballistic method, with the specimen in the form of a long wire, and a magnetizing coil 40 cm. long and 1.5 cm. diameter. The change in flux in a search coil of 1000 turns wound round the centre of the specimen, was measured by a Grassot fluxmeter (1.43×10^6 maxwell turns per radian).

The distribution curves are shown in figures 9, 11, 13. The number of counts per unit range of ΔM plotted as ordinates are reduced to number per cm³ in order to compare specimens of different sizes. In all cases the curves rise steeply near the noise level.

Size contribution curves are shown in figures 10, 12, and 14, giving in the case of annealed iron a reduction in gradient near the noise level. This reduction may be incorrect because of experimental errors which occur near the noise level, and the errors which are inevitable in differentiation.

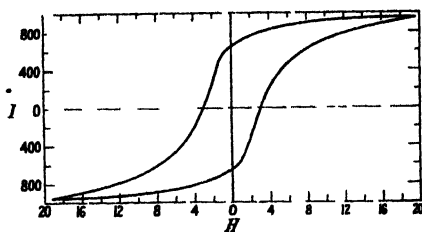


Figure 4. Hysteresis curve for annealed iron.

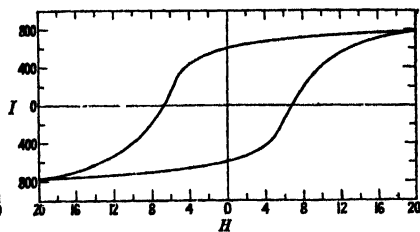


Figure 5. Hysteresis curve for hard-drawn iron.

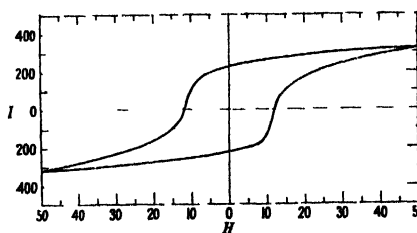


Figure 6. Hysteresis curve for hard-drawn nickel.

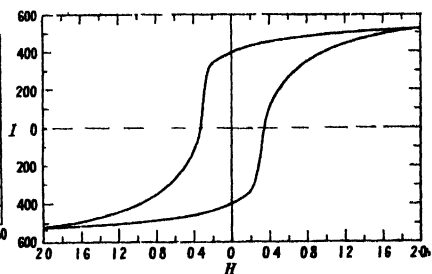


Figure 7. Hysteresis curve for hard-drawn mumetal.

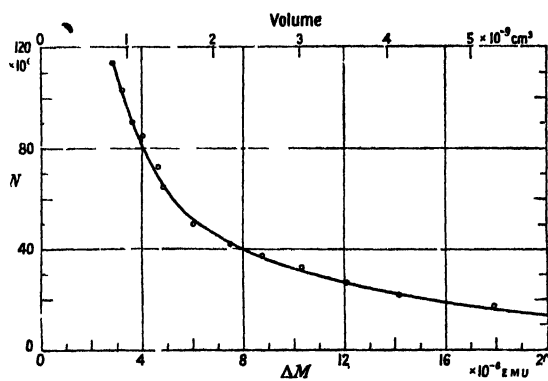


Figure 8. Number-level curve for annealed iron.

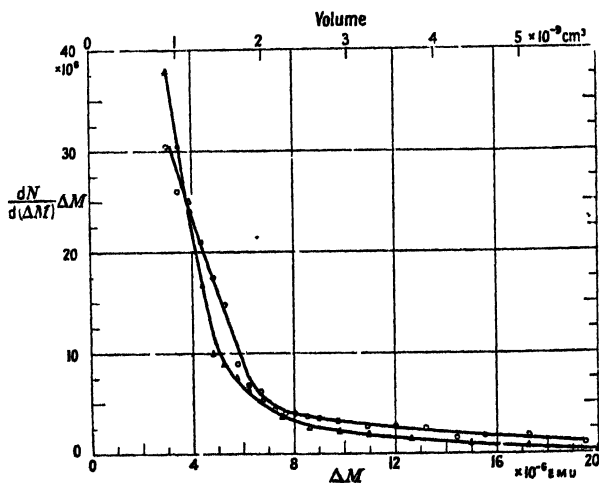


Figure 9. Distribution curves for annealed iron, O, and hard-drawn iron, Δ.

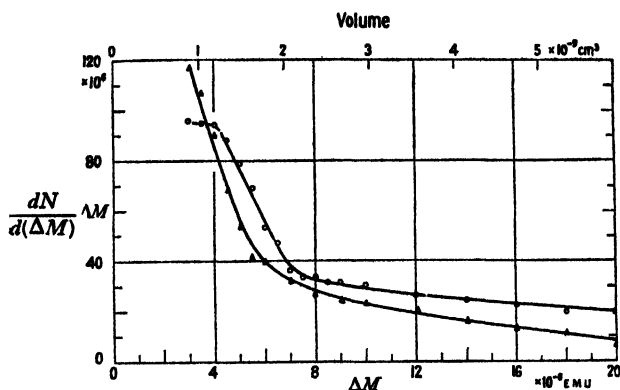


Figure 10. Contribution curves for annealed iron, O, and hard-drawn iron, Δ .

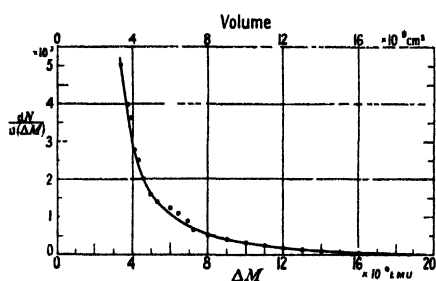


Figure 11. Distribution curve for hard-drawn nickel.

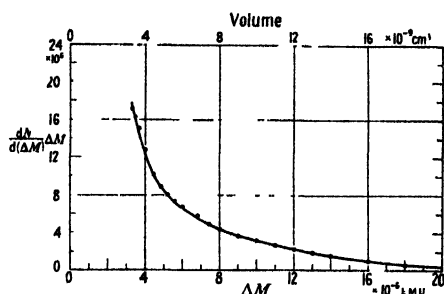


Figure 12. Contribution curve for hard-drawn nickel.

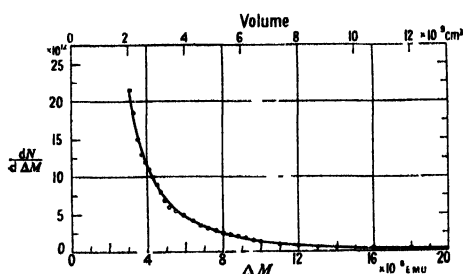


Figure 13. Distribution curve for hard-drawn mumetal.

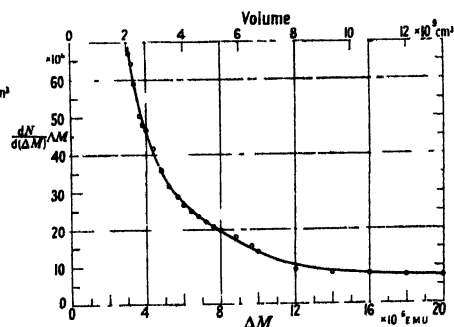


Figure 14. Contribution curve for hard-drawn mumetal.

The contributions to the total change in magnetization which are from the areas under the curves 10, 12, and 14, are given in table 1. The contribution of discontinuities lying above the noise level and below 20×10^{-6} E.M.U. are shown together with the extrapolated contributions. The extrapolated contributions were found in the following manner. The number-level curve over the range which could be examined is clearly of the exponential type. A reasonably satisfactory fit is obtained with a curve of the form $Ae^{-\alpha x^{\frac{1}{2}}}$. Differentiation of this gives the expression $\frac{1}{2}A\alpha x^{-\frac{1}{2}}e^{-\alpha x^{\frac{1}{2}}}$ and the contribution takes the form $\frac{1}{2}A\alpha x^{\frac{1}{2}}e^{-\alpha x^{\frac{1}{2}}}$.

Table 1. Change in magnetization contributed by Barkhausen discontinuities

Material	I_0	μ_0	Total change in I	Contribution of Barkhausen discontinuities				$\frac{dH}{dt}$ 10 ⁻³ oersted/sec.
				From 3 to 20×10^{-6} E.M.U.	% of total	Extrapolated contribution 0 to ∞	% of total	
Annealed iron	1700	140	1900	605	32	1300	68	3-15
Hard-drawn iron	1700	102	1540	475	31	840	55	3-15
Hard-drawn nickel	500	12.2	520	56	11	120	24	6-30
Hard-drawn mumetal	750	430	1040	320	31	520	50	0.15-1.5

The area under the contribution curve from $\Delta M = 0$ to $\Delta M = \infty$ is given by

$$\frac{1}{2} A \alpha x^{\frac{1}{2}} e^{-\alpha x^{\frac{1}{2}}} = 2A/\alpha^2,$$

which was taken as the extrapolated value of the total change in magnetization contributed by Barkhausen discontinuities.

As can be seen in the case of all the materials examined, the total change in magnetization is not accounted for by the Barkhausen discontinuities detected. If this comparison is restricted to the steep portion of the hysteresis cycle the agreement is no better. Further, the rate of change of the magnetizing field was varied within the limits shown, and this had no appreciable effect on the number of counts obtained. Some of the results for nickel, indicating this, are shown in table 2.

Table 2. Effect of variation of rate of change of field on number of counts obtained for hard-drawn nickel wire

Discriminator setting (ΔM) 10 ⁻⁶ E.M.U.	3.74	4.14	6.9
Time taken for one half-cycle (minutes)	19, 24, 24, 32, 39	23, 23, 25, 34, 40	27, 27, 30, 60
No. of counts in one half-cycle ($\times 10^{-3}$)	106, 119, 109, 103, 110	89, 90, 88, 88, 95	39, 40, 44, 40
Mean ($\times 10^{-3}$)	110 \pm 3	90 \pm 2	39 \pm 2

With the results obtained it was possible to show the variation in the number of discontinuities per unit range of magnetizing field dN'/dH , where $N' = dN/d\Delta M$ with field H , for different sizes of discontinuities. In order to display these results on a convenient scale, the contribution $\Delta M \cdot dN'/dH$ has been plotted against the field. These curves show maxima near the point $H = H_c$ where H

is the coercive field, as might be expected, and the contribution falls in the region where reversible changes predominate (figures 14, 15 and 16).

It should be possible to plot the number of discontinuities per unit change in intensity, dN'/dI , against intensity I for various sizes of discontinuity and thus obtain some information on the distribution of the discontinuities (number, size and contribution) as the intensity is altered over the hysteresis cycle. Unfortunately it is difficult to obtain a sufficient number of readings where the intensity is changing rapidly (near $H = H_c$).

The "average size" of domains detected can be obtained by dividing the total number of discontinuities counted into their contributions for different field ranges. The variation of this "average size" over the hysteresis cycle is shown for annealed iron in figure 18; this is a typical curve, showing a maximum near $H = H_c$. This curve is similar to those obtained by Bozorth and Dillinger. The significance of this curve is limited, as it only applies to those domains which are detected. If all the domains were accounted for the curve might be considerably different.

§ 6. DISCUSSION

For the materials examined it appears that discontinuous changes in magnetization in domains larger than 0.9 to 3.5×10^{-9} cm³ according to the material, do not account for the total change in intensity over the steep portion of the hysteresis curve. It would seem that at least some of the discrepancy is caused by domains of a size smaller than those detected by the present apparatus. The extrapolation of the experimental curves gives some estimate of the effect of these smaller discontinuities. Another factor which may account for some of the discrepancy is the extent to which reversible changes in magnetization take place over the hysteresis cycle. A quantitative estimation of this effect can be made by measurement of the reversible permeability at various points on the hysteresis cycle.

There are two ways of increasing the sensitivity of the apparatus so that smaller discontinuities can be detected: (a) the noise level of the amplifier could be reduced; (b) the sensitivity of the search coil could be increased.

In the case of (a) it is doubtful whether the valve noise can be further reduced to any appreciable extent. It does seem possible, however, that the search coils might be re-designed to give a greater sensitivity, without at the same time increasing the thermal agitation noise. This presents difficulties, as increasing the physical dimensions of the search coil will increase the effect of interference from external sources. An additional method might be to reduce

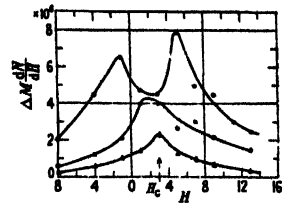


Figure 15. Contribution per unit change in H —annealed iron.

- $\Delta M = 3.2 \times 10^{-6}$ E.M.U.
- $\Delta M = 6.0 \times 10^{-6}$ E.M.U.
- △ $\Delta M = 21.0 \times 10^{-6}$ E.M.U.

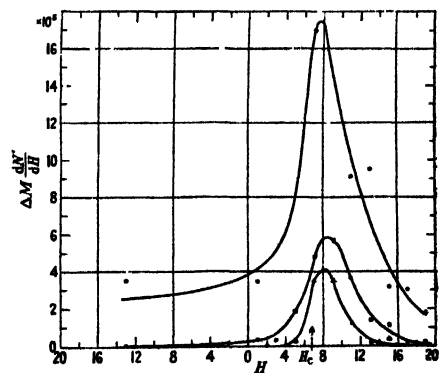


Figure 16. Contribution per unit change in H —hard-drawn iron.

- $\Delta M = 3.2 \times 10^{-6}$ E.M.U.
- $\Delta M = 6.0 \times 10^{-6}$ E.M.U.
- △ $\Delta M = 12.3 \times 10^{-6}$ E.M.U.

the temperature of the search coils, since the thermal agitation noise decreases with both resistance and temperature.

The effect of the rate of change of field could be investigated over a wider range, although an increase in this rate would appear to increase the possibility of coincidences in the discontinuities. The reduction in size of the specimen would reduce the effective rate of change of field by reducing the number of domains counted.

It is desirable that more results should be obtained near the region $H = H_c$, possibly by increasing the demagnetizing factor of the specimen, so as to reduce the effective gradient of the hysteresis curve in that region. Annealed mumetal would appear to be suitable for this purpose. It would be further interesting to investigate materials of the mumetal type, as it is possible that domain size bears some relation to initial permeability.

It is hoped to carry out experiments on materials covering a wider range of magnetic properties than has been possible so far.

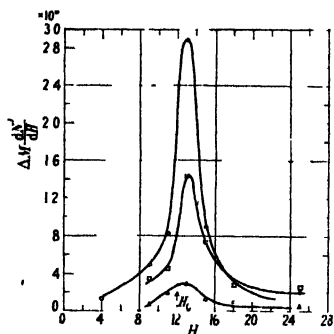


Figure 17. Contribution per unit change in H hard-drawn nickel

- $\Delta M - 3.9 \times 10^{-6}$ F.M.U.
- $\Delta M - 6.2 \times 10^{-6}$ F.M.U.
- Δ $\Delta M - 13.2 \times 10^{-6}$ F.M.U.

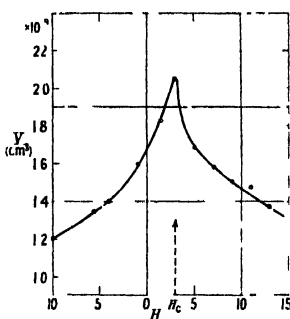


Figure 18. Average domain size for annealed iron.

ACKNOWLEDGMENT

The authors are much indebted to Professor R. Whiddington, under whose helpful direction this work was carried out, and to Professor E. C. Stoner, for his advice and generous assistance in the preparation of the manuscript. Thanks are due also to Messrs. J. Rigby and Sons Ltd., for supplying the iron specimens, to The Telegraph Construction and Maintenance Co. Ltd., who provided the mumetal, to the Admiralty for the loan of apparatus, to Mr. W. A. Ludbrooke, who constructed the counting recorder and search coils, and to Mr. H. Partridge through whose co-operation many of the components were obtained.

REFERENCES

- BARKHAUSEN, H., 1919, *Phys. Z.*, **20**, 401.
- BOZORTH, R. M., 1929, *Phys. Rev.*, **34**, 772.
- BOZORTH, R. M., and DILLINGER, J., 1930, *Phys. Rev.*, **35**, 733 ; 1932, *Ibid.*, **41**, 345.
- BUSH, H. D., 1947, Ph.D. Thesis, *The Barkhausen Effect*, Leeds University.
- LEWIS, W. B., 1942, *Electrical Counting* (Cambridge University Press).
- NEHER, H. V., 1939, *Rev. Sci. Instrum.*, **10**, 162.
- STEVENSON, E. C., and GETTING, I. A., 1937, *Rev. Sci. Instrum.*, **8**, 414.
- TYNDALL, E. P. T., 1924, *Phys. Rev.*, **24**, 439.

DISCUSSION

Prof. BATES. I should like to emphasize the importance of this work, which I very much hope will be extended. The trouble is that the Barkhausen effect is comparatively easy to show on the lecture table, and people have therefore been inclined to think that everything is known about it. Therefore it is only when a serious and careful study is made that we find, as the authors have clearly shown, things are not at all satisfactory. Moreover, there has been a tendency, particularly in the American and German literature, to give magnetic domains a kind of rigid cubic-block structure that they cannot possibly possess. I very much hope that the extension of this work will enable us to get a clearer view of domain processes, and, in particular, help us to decide on the correctness of the views concerning fillet-shaped domains which have recently been put forward.

Prof. W. SUCKSMITH. Do the authors believe that the discrepancy between the integrated sum of the intensity changes as determined by the Barkhausen discontinuities and the total change of intensity determined in the usual way is due to irreversible groups below the noise level, or to continuous and reversible changes?

Mr. McCAIG. I should like to point out that fast counting rates can now be measured without any special form of mechanical counter. Counting speeds of more than 100,000 per second have now been achieved in commercial scaling circuits.

AUTHORS' reply. Measurements on reversible permeability are being carried out at Leeds, the results of which suggest that reversible changes account for about 10% of the total change in magnetization over the range considered. It is possible that some of the discrepancy may be due to the assumption that the time constant of the discontinuity in the specimen is short compared with that of the search coil. This is based on a formula of Wwedensky (1921). It is intended to carry out measurements on the time constant of the discontinuities, making use of a wide-band amplifier (1 kc/s. to 1 Mc/s.) with which it should also be possible to detect the smaller discontinuities.

WWEDENSKY, 1921, *Ann. d. Phys.*, *Lpz.*, **64**, 609.

A Radiosonde Method for Atmospheric Potential Gradient Measurements

By R. E. BELIN

Dominion Physical Laboratory, Wellington, New Zealand

MS. received 9 July 1947

ABSTRACT. A method is outlined whereby the National Bureau of Standards radiosonde, in a modified form, is applied to the measurement of the earth's potential gradient. This paper describes the modification of the sonde circuit to measure the point-discharge current produced by a gradient and the result of a typical sounding of a cumulo-nimbus cloud.

§1. INTRODUCTION

EXPERIMENTS on the variation in natural potential gradient have so far been almost wholly carried out on or near the surface of the earth. The main exceptions were in the work of Simpson and Scrase (1927, 1935), who used a balloon to carry aloft, to heights of the order of 10 km., recording equipment capable of measuring pressure, humidity and potential gradient. The pressure and humidity were recorded, in the usual manner, by employing

an aneroid barometer and a human hair hygrometer respectively; the measurement and recording of the gradient was, however, more complex. Simpson and Scrase relied upon the fact that, if two sharp points with opposed senses are connected by a straight conductor, the axis of which is parallel to a field of suitable magnitude, point discharge will take place, thereby tending to equalize the potentials between the respective parallel planes through the points. In their apparatus this discharge current was caused to flow between two electrodes on polarity paper, and the magnitude of the gradient could be roughly estimated by measuring the width of the line traced by the positive electrode. Using a 20-metre length of fine copper wire to separate the points, it was possible to measure fields not smaller than 10v/cm. and the main disadvantage was that results depended upon the recovery of the equipment intact.

The method employed by the author (1946) differs essentially from that of Simpson and Scrase in that a permanent record is obtained at the receiving station so that results do not rely on the recovery of the sounding equipment. The point-discharge method was used in conjunction with a radiosonde instead of polarity paper. This was so modified that it was capable of transmitting a carrier, the modulation of which was dependent upon the magnitude of the electric field. These signals were detected and recorded by the standard radiosonde receiver developed by the Bureau of Standards (Diamond *et al.* 1940).

§2. METHOD

If two points be placed in an electric field as explained above, and a large-valued resistor be connected in series with the collector wire, at its centre, then the point-discharge current i flowing out of the points will be given (Simpson and Scrase 1927, 1935) by the relation

$$i = a(F^2 - M^2) \quad \dots\dots(1)$$

F is the existing potential gradient (greater than M). M is the minimum gradient required for point discharge to commence. a is a constant depending mainly on the exposure of the points (i.e. pointedness, number of points, distance between them etc.) and to a lesser degree upon the value of the series resistor. This current i produces across the resistor a potential drop which is measured and transmitted by the modified radiosonde.

Thus by carrying aloft (with a neoprene balloon) the collector points separated by a length L of wire with the modified transmitter at its centre the measurement of gradient can be readily achieved.

§3. MODIFIED RADIOSONDE

The radiosonde developed by Diamond, Hinman, Dunmore and Lapham (1940) consisted of two oscillators operating at frequencies of 1 Mc/s. and 70 Mc/s. respectively. The 1-Mc/s. oscillator M (figure 1) was of the squegger type. Its squegging frequency was determined by the R_2C_2 combination introduced into the grid circuit. The oscillator C was also of the squegger type being on when M is off and vice versa, thus effecting a modulation.

When M is oscillating it draws grid current which charges up the condenser C_2 and thus developing across R_2 a potential which builds up and eventually blocks the tube. The period of no oscillation is determined by the time constant R_2C_2 . If a bias battery P be introduced, as is done in figure 1, the squegging frequency will be altered. It is found that the squegging frequency has an

approximately linear relationship with the potential developed across R_2 (figure 2). This linear relation holds for positive bias, but only approximately for negative bias up to cut-off. Figure 2 gives steady grid voltage as a function of squegging frequency, when $R_2 = 2 \text{ M}\Omega$ and $C_2 = 0.003 \text{ }\mu\text{F}$.

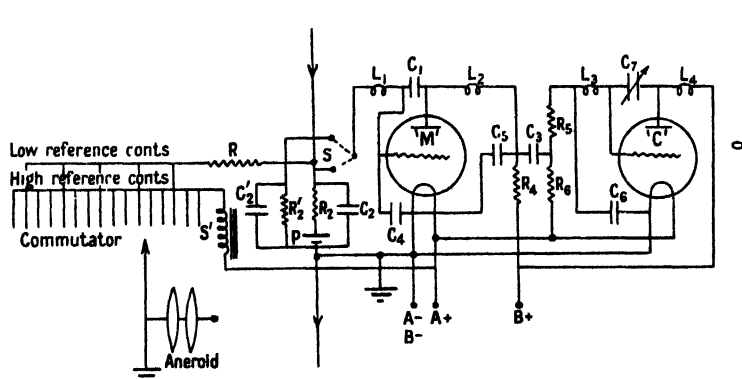


Figure 1. The modified radiosonde circuit.

$R = 1 \text{ M}\Omega$, $R_2 = 2 \text{ M}\Omega$, $R_1 = 1.25 \text{ M}\Omega$, $C_1 = 0.05 \text{ }\mu\text{F}$, $C_2 = 0.003 \text{ }\mu\text{F}$.

In figure 1 the small switch S , actuated by the standard relay S' on the 5, 10, 20, 25 etc. commutator contacts, was used to introduce periodically into the grid circuit of M a fixed time constant R_2' in parallel with C_2' . This periodicity was determined by the rate of change of pressure, or height, and hence was only controlled by the baro-switch. (The baro-switch consisted of a commutator composed of 80 separately insulated silver contacts of which only 5, 10, 15, 20 etc. were used. Over this commutator the contact arm of the aneroid barometer could move freely so that its position was determined by the pressure, that is by the height.) By means of this reference signal, as it was termed, it was possible to evaluate the altitude of the sounding equipment, as well as to obtain a check on any possible oscillator drift. The values of R_2' and C_2' were chosen to produce a squegging frequency of 190 c/s. and were used as the high reference value.

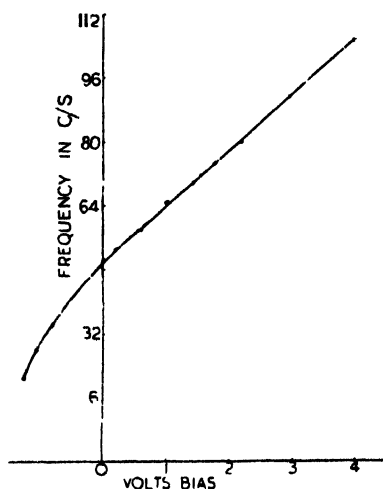


Figure 2.

The baro-switch also introduces on the 15, 30, 45 etc. contacts, a resistor R in parallel with R_2 . This latter combination produced, when no point discharge was taking place, a low reference of 140 c/s. Thus this low reference was subject to a change, the magnitude and direction of which would be dependent upon the magnitude and sign respectively of the potential gradient. However, the chief function of this reference was not so much a check on the oscillator as a check on the vertical course of the balloon. It served also as a height signal. It is possible should the ascent take place through a cumulo-nimbus cloud, as was intended, to encounter such a strong downward current as to produce a

descent of the balloon. This would produce a change in the predetermined sequence of high and low references which would be observed on the record.

The value of R_2 and C_2 were chosen to give an audio frequency of 50 c/s. with no bias battery in circuit. When the small $1\frac{1}{2}$ -volt penlight cell was introduced in series with R_2 , a squegging frequency of 64 c/s. was obtained. Thus the cut-off was moved 14 c/s. further to the left. This was an advantage should a large negative gradient occur.

§ 4. THE COLLECTORS AND RADIATING AFRIAL

Figure 3 shows a typical collecting device. It consists of pins soldered to the ends of the wires which form the framework. Two of these collectors with pins in opposed senses were used at either end of a gauge 30 copper connecting wire, at the centre of which was located the modified radiosonde. In order to

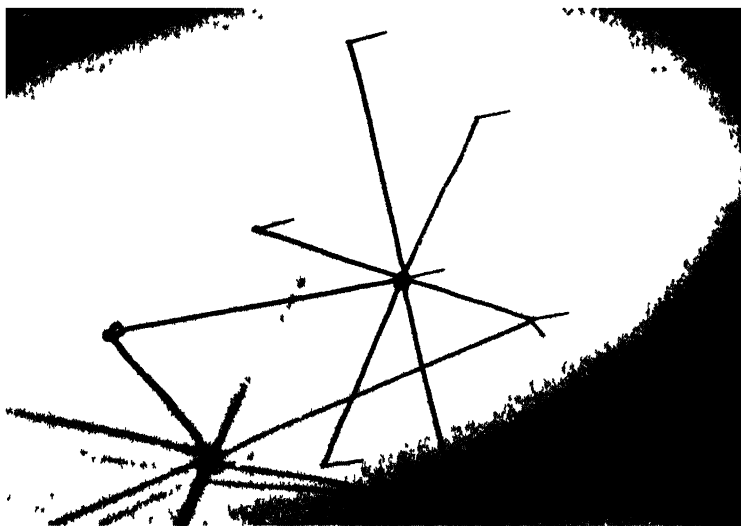


Figure 3. A typical collector, showing the manner in which the sharp points were connected.

support the sounding equipment which includes the collectors and radiosonde, the thin wire was wound round some strong hemp string, while the upper collector was fastened with a similar string, 10 feet long, to the balloon.

The transmitter aerial, a half-wave dipole, was fastened parallel to the collector wires, but separated from them by stand-off insulators of cork (for lightness) impregnated with polystyrene solution in toluene.

The length of the collector wire was chosen according to the formula $(2n - 1)$ half-wavelengths (where n is an integer) to give, as it was experimentally shown, a maximum signal strength. The value of n was chosen as 5 which thus gave a length of 64 feet between the collectors for a frequency of 69 Mc/s. The 64-foot connecting wire consisted of gauge 30 enamelled copper wire. For this length the minimum potential gradient capable of beginning point discharge was determined experimentally as 3V/cm. at N.T.P. Figure 1 shows schematically the collector wire connected to either end of the R_2C_2 combination, while the actual radiating antenna is seen on the right of the circuit.

§ 5. THE CALIBRATION

Calibration was carried out by means of an adjustable electrostatic field between two large parallel metal sheets. In this field were placed the collectors, actually used, with the pins parallel to the field direction. The collectors were connected by a short wire with the resistor R_2 in series, as the 64-foot length was impracticable.

The value of the potential between two parallel planes passing through the points was determined from the relation

$$\text{potential} = Ldv/dx \text{ volts,}$$

where L is the distance in cm. between the pin points and dv/dx the field strength in volts/cm. For points of given shape the value of Ldv/dx was shown to be constant for any value of L . Hence by calibrating this for a small given value of L , the calibration of frequency in cycles/sec. as a function of Ldv/dx was the same for any value of L^* . A calibration of a typical sonde using 7 points on both collectors is shown in figure 4. It will be observed that the minimum value of Ldv/dx necessary to initiate point discharge is approximately 6800 volts, corresponding, for the length of collector wire used, to 3.5 v/cm.; it is thus equal to the value of M required in equation (1), at the existing pressure, temperature and humidity.

The minimum value of Ldv/dx necessary to produce a discharge was shown to be independent of the number of points, but once discharge had set in, the point-discharge current became directly proportional to the number of points. Thus when the current for a given number of points is measured, that for any other number of points can be readily determined.

Point-discharge current is also a function of pressure related by the formula (Tamm 1901)

$$i = i_0 \{ P_0/P - (V/10^5)^4 \log_e(P_0/P) \}^2, \quad \dots\dots(2)$$

where i and i_0 are the currents at pressures P and P_0 respectively and $V = Ldv/dx$. P_0 is the pressure at ground level at time of releasing sounding equipment.

This cumbersome formula may be reduced, without affecting the accuracy required, to the approximate expression,

$$i = i_0 P_0^{1.6}/P, \quad \dots\dots(3)$$

which was used in these experiments.

The height of the sonde was determined from the formula

$$h = 8 \log_e P_0/P; \quad \dots\dots(4)$$

combining equations (3) and (4) we obtain

$$i = i_0 e^{0.2h}.$$

* By increasing the value of L it is possible to make the apparatus as sensitive as is desired.

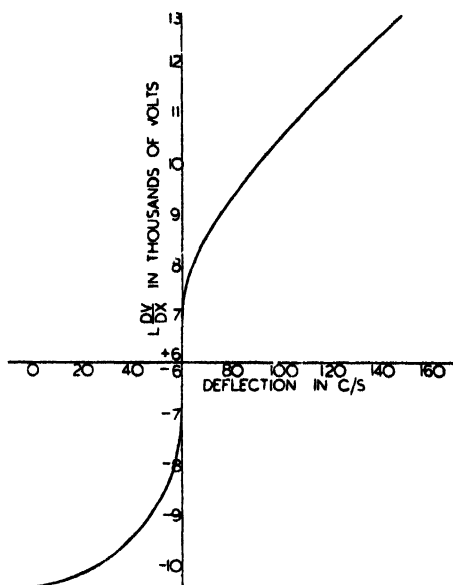


Figure 4.

Hence expression (1) can be modified to include a correction for height

$$i = a(F^2 - M^2)e^{0.2h}. \quad \dots\dots(5)$$

By the use of this formula the actual value of F , i.e. dv/dx , can be determined at any height.

§ 6. RESULTS

The equipment outlined above was used with only one modification, viz. thirteen points were used on the collector frames instead of seven. This altered the calibration by the factor 13/7. With the temperature, pressure and humidity prevailing at the time of the following ascent, the striking gradient value was observed to be 3 v/cm.; this is the value used in figures 5 and 6.

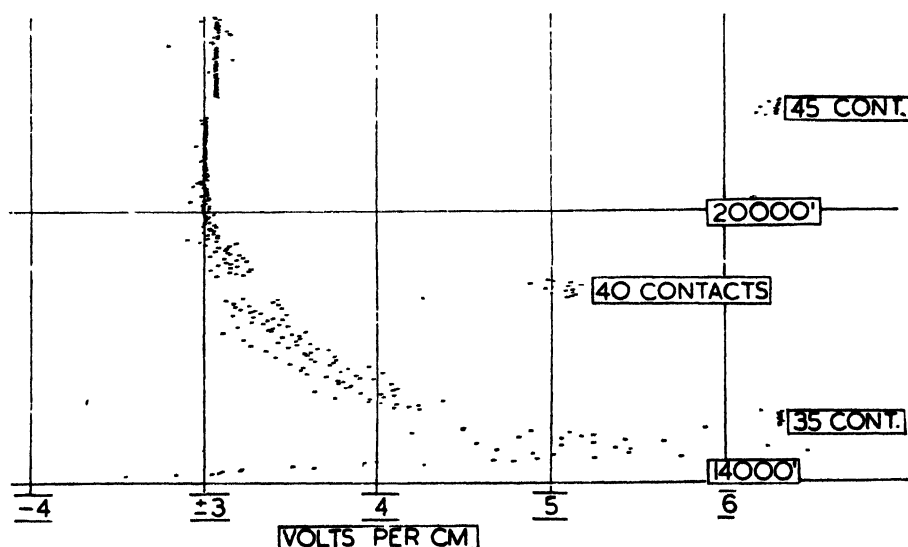


Figure 5.

Figure 6 shows a typical record obtained by releasing this apparatus beneath a cumulo-nimbus cloud with fracto-stratus formation below. The field at ground level was 4 v/cm. positive, that is, downwards. It increased in a fluctuating manner to a maximum of about 5 v/cm. at 600 feet, and then slowly decreased to zero at 2000 feet. The graph has been smoothed particularly in the lowest negative bulge between 2000 and 3500 feet which region probably corresponds to the fracto-stratus region. On entering the cloud at 3500 feet the potential gradient increased rapidly to a high positive value (off record) in the vicinity of 7000 feet. It then began to decrease again and at 10000 feet suddenly reversed to a high negative value (off record). At this stage the gradient was so large that the discharge current produced a voltage across the grid resistor R_g which was well beyond the cut-off value. However, a signal was still detected in the form of the 69-Mc/s. carrier wave. At 14000 feet, the cloud top, the gradient slowly became less negative, swinging over to positive when leaving the cloud, and finally reaching another subsidiary maximum of about 6 v/cm. at 14600 feet. Above this the gradient decreased again according to the inverse square law until at about 20000 feet the point discharge current became zero.

A reproduction of a part of the actual record obtained, in this case, when emerging from the cloud top is shown in figure 5, which corresponds to the upper part of figure 6. This record shows, by the scattered points, the fluctuating manne

in which the gradient increases or decreases during an ascent. This fluctuating increase could be explained by slight variations in the state of ionization with height, or possibly a humidity variation which might produce the same result.

Interpreting the curve of figure 6, we note that the cumulo-nimbus appeared to have a large positive distribution of electricity in the base and a large negative distribution in the top. This corresponds to a strong positive gradient below and above the cloud, the latter of which was observed. The former was distorted by the presence of the fracto-stratus formation on which it was assumed the main field of the cloud terminated, thus producing in effect a negative and a positive change in the upper and lower parts of the formation.

§7 CONCLUSION

The method outlined above is applicable to many problems relating to potential gradient: it is excellent, as proved, for the measurement of electrical distribution of clouds; it is possible, by lengthening L , to measure the fluctuation in dv/dx with height on a fine day, when it is approximately equal to $1v/cm$; by slight modification to the sonde it is possible to include in the record both humidity and temperature and thus obtain reliable quantitative data on the true variation of the potential gradient; again, by a further slight modification to the circuit which is at present being attempted, it will be possible to obtain measurements of very large gradients, especially negative, which have hitherto been beyond the range of the apparatus.

ACKNOWLEDGMENTS

The author is indebted to Dr. K. Kreielsheimer for valuable assistance and for the loan of his radiosonde receiver, without which these investigations could not have been undertaken, to Professor P. W. Burbidge for helpful suggestions, and to the R.N.Z.A.F. Meteorological Branch for materials, meteorological data and assistance with the balloon flights.

REFERENCES

- DIAMOND, HINMAN, DUNMORE, and LAPHAM, 1940, *Bur. Stand. J. Res.*, **25**, 327.
 KREIELSHEIMER and BELIN, 1946, *Nature, Lond.*, **157**, 227.
 SIMPSON and SCRASE, 1927, *Proc. Roy. Soc. A*, **114**, 376.
 SIMPSON and SCRASE, 1935, *Proc. Roy. Soc. A*, **161**, 309.
 TAMM, 1901, *Ann. Phys., Lpz.*, **6**, 259.

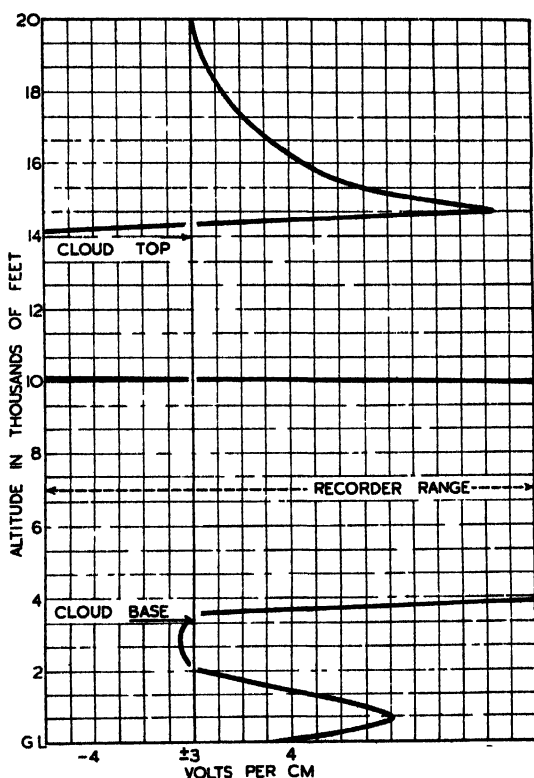


Figure 6.

A Note on the Lecher Wire Method of Measuring Impedance

By MARJORIE WILLIAMSON

Bedford College, London

MS. received 18 August 1947

ABSTRACT. Some time ago a new method of measuring impedance by means of a Lecher wire system was described by Dr. Williams in these Proceedings (1944). This method would seem to be very useful for the investigation of the properties of dielectrics since it gives both the real and the imaginary part of the dielectric constant, and it is also adaptable for use over a considerable range of frequencies.

In the course of some measurements on these lines at a wavelength of about a metre a number of interesting points have arisen.

BRIEFLY, in this method, Z , the impedance to be measured, is connected across one end of the Lecher wires which are loosely coupled to an oscillator at the other end. The standing-wave pattern is investigated by means of a pair of movable bridges consisting of vacuum thermojunctions of impedance Z_1 and Z_2 respectively, the values of which need not, however, be known.

In the position shown in figure 1 the currents passing through the bridges are I_1 and I_2 respectively. It has been shown (Williams 1944, equation (10)) that the modulus ρ of the ratio I_1/I_2 is given by

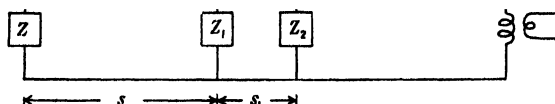


Figure 1.

$$\rho^2 = K_1 + \frac{K_2 + K_3 \sin 2(b + \beta s)}{\sinh^2 a + \sin^2(b + \beta s)} \quad \dots\dots (1)$$

where $\beta = 2\pi/\lambda$, the current reflection coefficient

$$(Z_0 - Z)/(Z_0 + Z) = e^{-2(a + j\beta s)}, \quad \dots\dots (2)$$

$$K_1 = (1/|Z_1|^2)[(R_2^2 + X_2^2) \cos 2\beta s_1 + X_2 Z_0 \sin 2\beta s_1 + Z_0^2 \sin^2 \beta s_1], \quad \dots\dots (3)$$

$$K_2 = (1/|Z_1|^2)[(R_2^2 + X_2^2) \sin^2 \beta s_1 \cosh 2a + R_2 Z_0 \sin^2 \beta s_1 \sinh 2a], \quad \dots\dots (4)$$

$$K_3 = (1/|Z_1|^2)[(R_2^2 + X_2^2) \sin \beta s_1 \cos \beta s_1 + X_2 Z_0 \sin^2 \beta s_1] \quad \dots\dots (5)$$

and Z_0 is the characteristic impedance of the Lecher wires.

In order to simplify equation (1) the constant K_3 is made zero by adjusting the value of s_1 . The necessary condition is, from (5),

$$(Z_2^2 + Z_0 X_2 \tan \beta s_1) \sin \beta s_1 \cos \beta s_1 = 0,$$

i.e.

$$\tan \beta s_1 = -Z_2^2/Z_0 X_2, \quad \dots\dots (6)$$

the other solutions being of no practical importance.

This critical value of s_1 , which we shall call s_c , was determined experimentally in the original method. The impedance Z was removed and the lines short

circuited by a copper plate. In this case a and b should be zero and the equation (1) reduces to

$$\rho^2 = K_1 + K_2 \operatorname{cosec}^2 \beta s + 2K_3 \cot \beta s, \quad (7)$$

K_1 and K_3 being unchanged and K_2 simplified by the vanishing of the term with $\sinh 2a$.

If now $K_3 = 0$, a graph of ρ^2 against s is symmetrical about its turning points. A finite value of K_3 introduces an asymmetry and this, in the original paper, is the test for the critical setting.

For $K_3 = 0$ a graph of ρ^2 against $\operatorname{cosec}^2 \beta s$ is a straight line from which K_1 , which is necessary for the determination of a , is obtained.

It was noticed, however, in our experiments, that to a (ρ^2, s) curve which appeared to be symmetrical there corresponded a $(\rho^2, \operatorname{cosec}^2 \beta s)$ graph which was a curve with two branches as shown in figure 2, i.e. K_3 was still finite, and this suggested that there might be a more sensitive test of the accuracy of the setting. It was thought of interest to examine the equation (7) more closely for values of s_1 not equal to s_c in the hope of finding a quick and accurate way of determining s_c .

The form of the curve given by equation (7) is shown in figure 3 which is a graph of the simplified function

$$\rho^2 = \operatorname{cosec}^2 \theta + K \cot \theta$$

from $\theta = 0$ to $\theta = \pi$ for the arbitrary values $K = 1$, $K = \frac{1}{2}$. For values of θ greater than π (i.e. for increasing s) the curve is repeated.

As θ increases the curve is traversed in the direction shown by the arrows if K is positive, and in the reverse direction if K is negative. Hence from the experimental curves it is quite obvious when the critical setting has been overshoot. As s_1 approaches the value s_c , i.e. as K tends to zero, the curves close in and approximate more and more nearly to the required straight line.

A number of experimental curves were obtained for various values of s_1 , and while the general behaviour corresponded to the above analysis in that the branches of the curves closed in and opened out again on passing through the critical setting, most of the curves showed unexpected peculiarities of curvature,

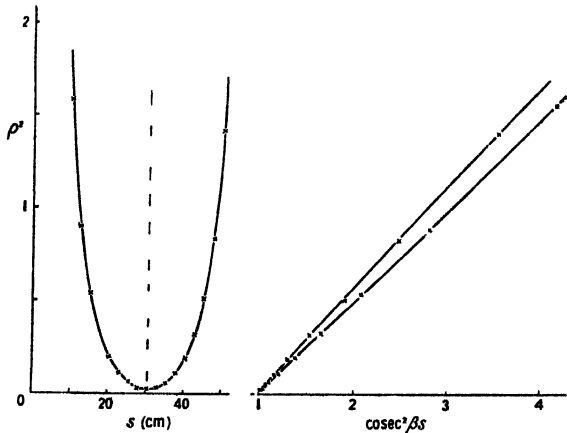


Figure 2.

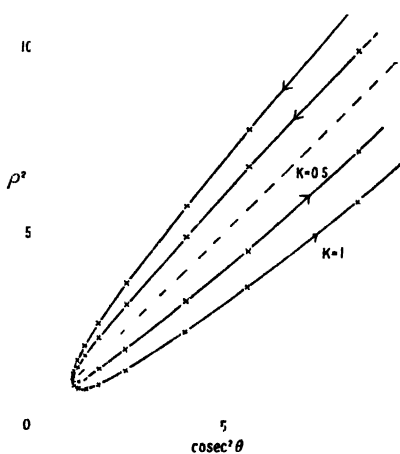


Figure 3.

and in some cases there was a crossing of the two branches. In no case could a reasonably good straight line be obtained.

There is nothing in equation (7) to account for this, but a consideration of the general equation (1) shows that if a or b or both are finite, though small, the $(\rho^2, \operatorname{cosec}^2 \beta s)$ graph is considerably altered.

Take first the case $a=0$, b finite; then

$$\rho^2 = K_1 + K_2 \operatorname{cosec}^2(b + \beta s) + 2K_3 \cot(b + \beta s).$$

A graph of ρ^2 plotted against $\operatorname{cosec}^2(b + \beta s)$ would give the curve of figure 3, but experimentally the graph plotted is of ρ^2 against $\operatorname{cosec}^2 \beta s$, and this shows a crossing of the two branches at some point, since, at any particular value of ρ^2 , the value of βs is too small.

Figure 4 is a graph of the simplified function

$$\rho^2 = \operatorname{cosec}^2(\theta + 10^\circ) + K \cot(\theta + 10^\circ)$$

plotted against $\operatorname{cosec}^2 \theta$ and should be compared with figure 3.

The point of crossing over depends upon the value of b and comes nearer to the bend of the curve as b increases. The width of the loop, however, that is the value of K , also affects the position of this point, so that it does not seem possible, as was hoped at first, to get any information about a or b from the form of this curve.

Considering now the case a finite, $b=0$, which gives

$$\rho^2(1 + \sinh^2 a \operatorname{cosec}^2 \beta s) = K_1 + (K_1 \sinh^2 a + K_2) \operatorname{cosec}^2 \beta s + 2K_3 \cot \beta s,$$

the general effect of the finite value of a can be seen.

Owing to the multiplying factor on the left-hand side which is always greater than unity, the ordinates will be reduced by a fraction which increases with $\operatorname{cosec}^2 \beta s$ so that there will be a downward curvature superimposed upon both branches of the curve. At the same time there will be a change in the slope of the whole graph due to the change in the constant which multiplies $\operatorname{cosec}^2 \beta s$ on the right-hand side.

Figure 5 shows the general curve, plotted for arbitrary values of the constants with both a and b finite. Figure 6 is one of the experimental curves for comparison.

It was concluded then that the copper plate was not a sufficiently good short circuit, and an estimate of the values of the constants a and b was made from the nearest approach to a straight line. These values were not large.

A copper plate of larger size was tried and gave a slight improvement, but evidently the attachment of the plate to the lines was contributing to the reflection coefficient and this quantity could not be reduced to zero in our experiments.

In measuring the reflection coefficient of an unknown impedance Z the quantity b is obtained from the displacement of the minimum of the (ρ^2, s) curve when the short circuit is replaced by Z . To find a a straight line plot $1/(\rho^2 - K_1)$

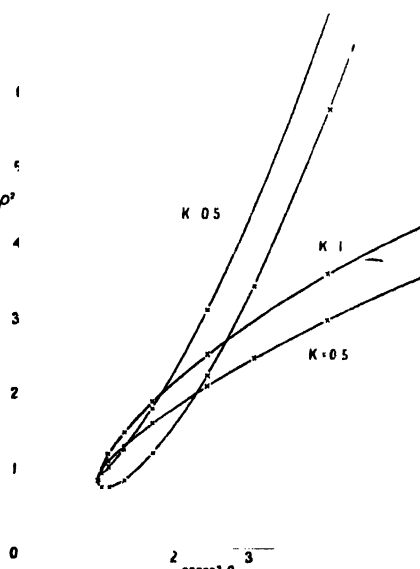


Figure 4.

against $\operatorname{cosec}^2 \beta s$ is required and this can only be obtained at the critical separation. Hence, if the lines cannot be satisfactorily short circuited, the critical separation is not accurately obtainable, and there is a certain amount of error in the determination of the reflection coefficient. The error is likely to be greater in a than in b , since the latter quantity is found more directly. Unfortunately it is just the quantity a which is the important one in the case of those dielectrics which show absorption.

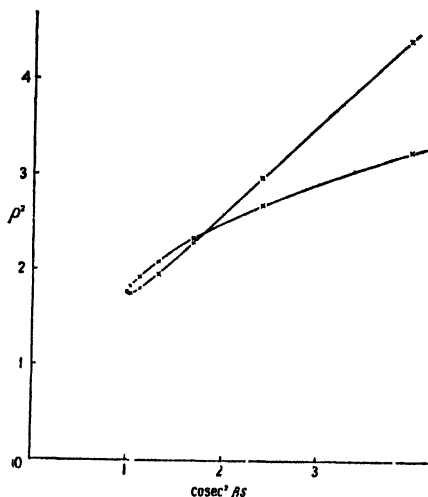


Figure 5.

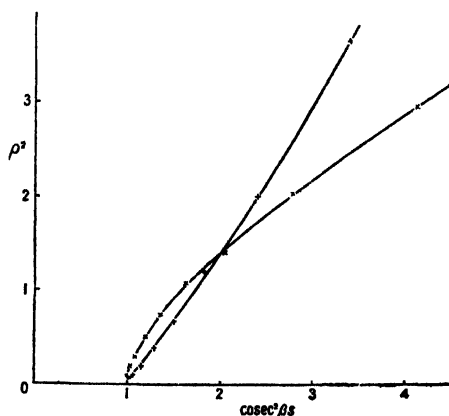


Figure 6.

The experimental work to test these points has been carried out by Miss E. B. Harriss and forms part of the thesis submitted by her for the M.Sc. degree of the University of London. The author wishes to thank her for her help, and also to thank Professor Flint, with whom the problems have been discussed.

REFERENCE

WILLIAMS, G., 1944, *Proc. Phys. Soc.*, **56**, 63.

Slip at Grain Boundaries and Grain Growth in Metals

By N. F. MOTT,
H. H. Wills Physical Laboratory, Bristol

MS. received 6 January 1948

§ 1. INTRODUCTION

IN a recent paper from the Chicago Institute for the Study of Metals, Ting-Sai-Ke (1947) gives an expression for the rate of slip at the grain boundaries in pure polycrystalline aluminium; this is derived from measurements of the internal friction and some other anelastic phenomena. The expression obtained for the velocity of slip at each boundary is

$$v = A\sigma e^{-B/RT} \text{ cm/sec.}, \quad \dots\dots(1)$$

where σ is the stress applied at the boundary in dynes/cm², $B = 34.5 \text{ kcal/gm. atom}$, and $A \simeq 18 \text{ (cm/sec.)/(dyne/cm.)}$.

Ke also points out the following fact : let it be assumed for purposes of calculation that between the grains there is a layer of fluid of thickness t and viscosity η . Then the velocity v of flow should be given by the formula

$$v = \sigma t / \eta. \quad \dots (2)$$

Suppose then that we extrapolate the formula (1) deduced from experiment to the melting point of aluminium, and then compare the results with (2), taking for η the viscosity of liquid aluminium at the melting point. We may deduce a value for t . This comes out to be a few angströms only. This suggests, as is probable on other grounds, that the boundary is not a thick amorphous layer, but merely the surface of contact of two crystals fitting together as well as their different orientations allow.

In this note we shall put forward a theory of the slip process and, in particular, of the factors which determine the constants A and B .

§ 2. SLIDING OF ONE SURFACE OVER ANOTHER

Let us consider two close-packed surfaces of metal in contact and with the same orientation, and determine the activation energy required to slide one over the other. This quantity is required as a preliminary to the subsequent calculations.

If the distance between the planes is h , the stress σ required to produce a displacement x is given by

$$\sigma = Gx/h,$$

where G is the shear modulus for the direction of shear considered. This expression is valid only if x is small. For large x we may set as an approximation

$$\sigma = (Ga/2\pi h) \sin(2\pi x/a),$$

where a is the distance to the nearest stable position of equilibrium; the use of the sine function gives an upper limit to σ . Thus the activation energy per unit area is, to this approximation,

$$\sigma dx = Ga^2/2\pi^2h.$$

If $2d$ is the lattice constant, then we may set

$$d = a, \quad a = 2d/\sqrt{3},$$

and the number $(1/\omega)$ of atoms per unit area is $1/\sqrt{3}d^2$. Therefore the activation energy nU required to slide n atoms of the top plane over the lower plane is

$$nU \sim 3nGa^3/4\pi^2.$$

For aluminium $a = 2.02 \times 10^{-8}$ cm., and since the metal is isotropic we may set for G the shear modulus of the bulk material :

$$G = 2.5 \times 10^{11} \text{ dyne/cm.}$$

This gives

$$U = 0.066 \text{ ev.} = c. 1.5 \text{ kcal.}$$

The smallness of this value is to be noted; it is much smaller than the heat of vaporization (76 kcal.), and of the order of the heat of fusion (2.55 kcal.).

§ 3. THEORY OF INTERCRYSTALLINE SLIP

If two crystal planes are in contact, but cannot fit owing to different indices or orientation, one may suppose that the surface of contact is divided into islands

where the fit is reasonably good, separated by lines near which fit is bad. We may suppose that on one of these islands the top plane can move relatively to the one underneath, and that this is the elementary act responsible for intercrystalline slip. With this hypothesis the rate of slip might be estimated as follows :

Let each island contain n atoms and thus cover an area $n\omega$. In the absence of stress the activation energy for slip is nU . In the presence of a stress σ it is $nU \pm \frac{1}{2}\sigma an\omega$, it being supposed that $\frac{1}{2}a$ is the distance to the potential barrier separating one position of equilibrium from the next. Thus the number of times per second that each island moves a distance a in the direction of the stress is

$$\nu \exp \{ -(nU - \sigma an\omega)/kT \},$$

and in the opposite direction

$$\nu \exp \{ -(nU + \sigma an\omega)/kT \},$$

where ν is the frequency of atomic vibration. It follows that the rate of slip is

$$\nu e^{-nU/kT} \sinh(\sigma an\omega/kT).$$

If σ is not too great this becomes

$$(\nu a^2 n \omega \sigma / kT) e^{-nU/kT}.$$

If nU is to be identified with the observed 34.5 kcal., we must have

$$n = 34.5/1.5 \simeq 23.$$

With $\nu \simeq 10^{12} \text{ sec}^{-1}$, $a \simeq 2 \times 10^{-8} \text{ cm.}$, $\omega \simeq 7 \times 10^{-16} \text{ cm}^2$, $T \simeq 600$, we get for A

$$A = \nu a^2 n \omega / kT \simeq 6 \times 10^{-5}.$$

In this treatment it has been assumed that σ at the point of slip is equal to the applied stress. Some stress magnification is possible, in which case A would be somewhat greater.

The objections to this treatment are :

(a) The calculated value of A is too small by a factor 10^4 – 10^5 .

(b) The observed fact that at the melting point the slip is the same as that which would be given by a monomolecular layer of liquid appears in the theory as an *accident*. The mechanism of flow in liquid aluminium cannot be anything like that sketched here, because the temperature dependence is 10–20 times smaller; the viscosity of liquid metals depends on temperature according to the formula

$$\sigma = \sigma_0 e^{W/kT},$$

where W is of the order of the latent heat of fusion (Frenkel 1946).

We must therefore modify our hypothesis. Let us suppose that the elementary act which allows slip to occur is the *disordering* of atoms round each island where fit is good. The free energy F necessary to do this will approach zero at the melting point and nL at the absolute zero of temperature; here L is the latent heat of fusion per atom. At any other temperature, let us assume F to be given by

$$F = nL(1 - T/T_M),$$

where T_M is the temperature of melting. Let us assume also, since the disordering will result in a slip through a distance a , that a stress σ will decrease or increase F by $\pm \frac{1}{2}\sigma n\omega a$. Then the rate of slip is now

$$v = 2\nu a \exp \{ -nL(1 - T/T_M)/kT \} \sinh(\sigma n\omega a/2kT),$$

which for small σ reduces to

$$v = \frac{\nu a^2 n \omega \sigma}{kT} \exp \left(\frac{nL}{kT_M} \right) \exp \left(\frac{-nL}{kT} \right).$$

For aluminium the latent heat L is 2.55 kcal/mol, so that n (the number of atoms to be disordered) should be about 14.

The factor $\exp(nL/kT_M)$ is $\exp(34\,000/2 \times 930) \simeq 10^8$. One is thus able to explain the large factor outside the exponential. One also sees why the surface layer behaves like a layer of liquid at the melting point, since the activation energy for melting small volumes there disappears.

§ 4. VELOCITY OF RECRYSTALLIZATION

Anderson and Mehl (1945) have shown that the velocity of growth of new crystals during recrystallization of cold-rolled aluminium is given by

$$v = v_0 e^{-Q/RT},$$

where $Q \simeq 55$ kcal. and $v_0 \simeq 10^{14}$ – 10^{16} cm/sec. Burgers and the present author (1947) have suggested that this very high value of v_0 indicates that some process which depends on temperature through the factor $e^{-Q/RT}$ triggers off the change of crystal form of a whole mosaic block, which then changes its crystal structure to that of the growing crystal without the intervention of any other thermally activated process.

An alternative and perhaps more plausible explanation can be given along the lines of the last section. Let us assume that the crystal grows at each step by the disordering or melting of a group of n atoms of the cold-worked material which then takes up the structure of the growing crystal. We have at present no theoretical means of determining n ; as in the previous problem it must be the value for which the free energy of disordering, including surface energy, is a minimum.

We should thus expect as in the last section a rate of crystal growth

$$va \exp(Q/RT_M) \exp(-Q/RT). \quad \dots\dots(3)$$

With $va \simeq 10^{13}$ cm/sec., and the first exponential factor equal to 10^{18} , this gives $10^{31} e^{-Q/RT}$, which is of the observed order of magnitude.

This analysis assumes that the local melting always results in an addition to the growing crystal and never the reverse process. To test this assumption, let q be the mean energy of cold work per atom. The condition for the validity of the assumption is then $nq \gg kT$. Otherwise (3) should be multiplied by $(1 - e^{nq/kT})$. For heavily cold-worked aluminium according to Taylor and Quinney (1934) q is 30 cal/mol. With $n \simeq 20$ this gives nq/kT at 600° K. of the order $\frac{1}{2}$. For lightly rolled aluminium the value should be considerably less, but the order of magnitude of (3) will not be affected.

REFERENCES

- ANDERSON, W. A., and MEHL, R. F., 1945, *Amer. Inst. Min. and Met. Engng.*, Tech. Pub. No. 1805.
 BARRETT, C. S., 1943, *Structure of Metals*.
 BURGERS, W. G., 1947, *K. Ned. Ak. Wet.*, **50**, 452.
 FRENKEL, J., 1946, *Theory of Liquids* (Oxford: Clarendon Press).
 TAYLOR, G. I., and QUINNEY, H., 1934, *Proc. Roy. Soc. A*, **143**, 307.
 TING-SUI-KE, 1947, *Phys. Rev.*, **71**, 533.

LETTERS TO THE EDITOR

Paramagnetic Resonance at Low Temperatures in Chromic Alum

The use of centimetre wavelengths to determine the small splitting (due to the crystalline electric field) of the basic spin quadruplet of Cr^{++} in potassium chromic alum has been described by Bagguley and Griffiths (1947), who find a value of 0.12 cm^{-1} for the separation between the $\pm 1/2$ and $\pm 3/2$ levels. The same authors (unpublished) have obtained the value of 0.13 cm^{-1} for the analogous splitting in ammonium chromic alum; independently Weiss, Whitmer, Torrey and Jen-Sen Hsiang (1947) give a value $0.15 \pm 0.01 \text{ cm}^{-1}$ for this salt, obtained by a similar method. These values differ considerably from those obtained at low temperatures by other methods: viz. $0.16-0.19 \text{ cm}^{-1}$ for potassium chromic alum (Bleaney 1939, Du Pré 1940, Casimir, de Haas and de Klerk 1939, Starr 1941, Broer 1947, Gorter *et al.* 1940, 1942) and 0.34 cm^{-1} for ammonium chromic alum (Starr 1941). The application of centimetre wavelengths to the determination of the change of splitting with temperature was therefore undertaken by the authors.

A fixed tuned cavity is used, resonant in the H_1 mode at about 3 cm. wavelength; it is excited by probes at the ends of small concentric line feeders. The cavity is surrounded by a dewar vessel, containing liquid oxygen, liquid hydrogen etc.; its temperature is determined by a vapour pressure thermometer, or by a thermocouple calibrated at various fixed points.

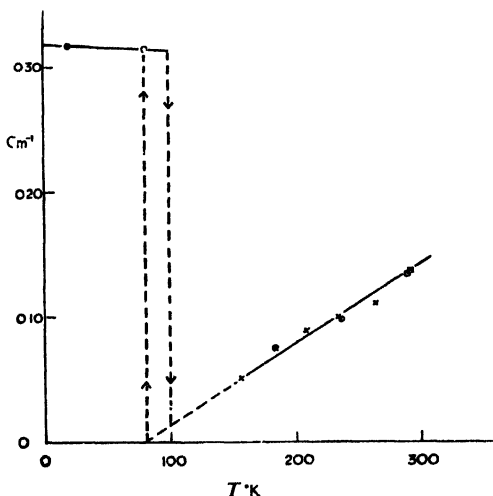
The magnification factor Q of the empty resonator is about 13 000 at 90° K. and 15 000 at 20° K. The magnetic absorption in the crystal rises with $1/T$, and it is therefore possible to obtain large effects from very small crystals. The crystal is placed in the resonator in a position of maximum magnetic field and minimum electric field (to avoid dielectric loss). The power transmitted through the cavity at resonance is observed by means of a silicon-galena rectifier while a variable magnetic field is applied, perpendicular to the R.F. magnetic field.

As the temperature is lowered, the absorption curve for ammonium chromic alum remains similar to that observed at room temperature, except that the side peaks due to the Stark splitting move steadily inwards towards the main peak, showing that the splitting is decreasing. At about 80° K. , however, a transition sets in, which requires several minutes for its completion. It is accompanied by a considerable change in the resonant frequency of the cavity, and the absorption curve obtained on applying a magnetic field after the transition is complete is quite different from that obtained at higher temperatures. On warming up slowly, the reverse transition does not occur until about 100° K. , after which the magnetic absorption curve reverts to that previously obtained at this temperature. These transitions are repeatable, but the exact transition temperature varies by a few degrees; in general the hysteresis between the transition temperatures for cooling and warming decreases after the operation has been repeated a number of times.

There is little doubt that this effect is the same as that found by Guillien (1939), who discovered a discontinuity in the dielectric constant, accompanied by jumps in the dielectric loss and specific heat, showing temperature hysteresis. The change in resonant frequency of the cavity observed by us is consistent in sign with this dielectric constant change, though comparatively small since the crystal projects only slightly into the R.F. electric field. The effects are attributed by Guillien to "co-operative relaxation of librational restraint on water or water-containing complexes with rising temperature". If the water molecules involved are those forming the octahedron around the Cr^{++} ion, the great influence of this transition on the Stark splitting of this ion can be understood, since Van Vleck (1939) has shown that the splitting can be explained by the slight distortion of the octahedron, together with other contributions due to the Jahn-Teller effect and to more distant ions. The transition has also been observed by Kraus and Nutting (1941), who found that the crystal often became practically opaque, as though shattered into an enormous number of tiny crystals; this is consistent with our observations. This shattering process does not, however, alter the magnetic absorption spectrum observed at higher temperatures.

Below the transition point, the magnetic absorption curve observed with the external magnetic field along the normals to the (111), (110), (100) crystal planes is quite inconsistent with that to be expected from the theory of Broer (1942), assuming a small trigonal field superposed on the cubic field. Since the results at room temperature are in good agreement with this theory, one must assume that at low temperatures the crystalline electric field is differently aligned or has different symmetry. The quantitative interpretation of the magnetic absorption curve is therefore not possible at this stage. In this salt, however, there is considerable absorption at 3 cm. wavelength in zero magnetic field below the transition point, showing that the splitting is of this order. By examining the absorption curve in low fields, the splitting has been determined as 0.314 cm^{-1} just below the transition point, increasing slightly to 0.317 cm^{-1} at 20°K .; the probable error in these values is less than 1%. Since, by Kramers' theorem, an electric field cannot do more than decompose the spin quadruplet into two doublets, these values for the position of maximum absorption in zero field must correspond to the Stark splitting, irrespective of the nature of the crystalline field.

The variation of the Stark splitting down to 20°K . is shown in the figure for ammonium chromic alum. Above the transition point the splitting varies linearly with temperature, and is, presumably, associated with rotation of the water molecules. Below the transition point, the slight change of the splitting may be due to thermal contraction, and it seems unlikely that there will be any substantial change below 20°K . The results of preliminary measurements on potassium chromic alum show that there is a similar decrease in the splitting on cooling downwards from room temperature, but no sudden transition occurs. Instead, abnormally weak side peaks appear in the absorption curve which, though not consistent with a trigonal field directed along the space (body) diagonal of unit cell, appear to correspond to a splitting of the same order as that obtained by other methods at low temperatures. Measurements are in progress at longer wavelengths to determine this splitting directly by a method similar to that used for the ammonium alum.



Variation of Stark splitting in ammonium chrome alum with temperature.

- Bagguley and Griffiths.
- × Before shattering.
- After shattering.

The Clarendon Laboratory,
Oxford.

19 December 1947.

B. BLEANEY.
R. P. PENROSE.

BAGGULEY, D. M. S., and GRIFFITHS, J. H. E., 1947, *Nature, Lond.*, **160**, 532.

BLEANEY, B., 1939, unpublished, quoted by Van Vleck, 1940.

BROER, I. J. F., 1942, *Physica*, **9**, 547; 1947, *Ibid.*, **13**, 353.

CASIMIR, H. B. G., DE HAAS, W. J., and DE KLERK, D., 1939, *Physica*, **6**, 365.

DIJKSTRA, L. J., GORTER, C. J., and VAN PAEMEL, O., 1942, *Physica*, **9**, 673.

DUPRÉ, F. K., 1940, *Physica*, **7**, 79.

GORTER, C. J., DIJKSTRA, L. J., and GROENDIJK, H., 1940, *Physica*, **7**, 625.

GUILLIEN, R., 1939, *C.R. Acad. Sci., Paris*, **209**, 21.

KRAUS, D. L., and NUTTING, G. C., 1941, *J. Chem. Phys.*, **9**, 133.

STARR, C., 1941, *Phys. Rev.*, **60**, 241.

VAN VLECK, J. H., 1939, *J. Chem. Phys.*, **7**, 61; 1940, *Phys. Rev.*, **57**, 426.

WHEISS, P. R., WHITMER, C. A., TORREY, H. C., and JEN-SEN HSIANG, 1947, *Phys. Rev.*, **72**, 975.

Milli-microsecond Pulse Generation by Electron Bunching

Work is in progress in this laboratory on a project for coincidence counting of nuclear particles with a resolution of the order of 10^{-9} seconds. For a study of transients of this duration, the first requirement is a technique of milli-microsecond pulse generation; these periods of time are below the limit of normal electrical pulse production methods. Equipment has been built for the production of these pulses by bunching electrons with alternating voltages of frequency 210 Mc/s., on the assumption that the mathematical theory of electron bunching applies (Webster 1939, Harrison 1947).

The figure shows in diagram the electronic arrangement, which is in effect a klystron with the resonant catcher replaced by a non-resonant collector in the form of a coaxial line. Power from a 210-Mc/s. oscillator is fed into a coaxial line resonator short-circuited at one end and capacitively terminated at the other by two grids 1 cm. apart, across which R.F. potentials up to 400 volts can be developed. An electron beam, after modulation by these grids, is magnetically focused on to the end of a 70-ohm coaxial line a distance $s = 40$ cm. away. The outer conductor is terminated in a grid through which the beam passes before reaching the inner conductor 4 mm. away. The bunched beam passing across the capacity between the beam and the conductor induces a voltage pulse which is then propagated down the line. The line is terminated in a probe, which feeds a length of 10-cm. waveguide through a "door knob" transformer, and a milliammeter by which the D.C. beam current, of the order of 2 ma., is read. The waveguide is terminated by a 10-cm. mixer with local oscillator, and a 60-Mc/s. I.F. amplifier with diode current meter (Collie, Hasted and Ritson 1948). This serves as a narrow band power measuring device with which it is possible to pick out a harmonic component of the pulse. The peak alternating voltage in the bunching pulse generator, that is the peak energy increments given to the electrons, can be measured directly by back-biasing the electrode on which the beam impinges until no current is observed either with or without the alternating voltage at the buncher grids.

The bunched beam passing across the capacity between the beam and the conductor induces a voltage pulse which is then propagated down the line. The line is terminated in a probe, which feeds a length of 10-cm. waveguide through a "door knob" transformer, and a milliammeter by which the D.C. beam current, of the order of 2 ma., is read. The waveguide is terminated by a 10-cm. mixer with local oscillator, and a 60-Mc/s. I.F. amplifier with diode current meter (Collie, Hasted and Ritson 1948). This serves as a narrow band power measuring device with which it is possible to pick out a harmonic component of the pulse. The peak alternating voltage in the bunching pulse generator, that is the peak energy increments given to the electrons, can be measured directly by back-biasing the electrode on which the beam impinges until no current is observed either with or without the alternating voltage at the buncher grids.

The theory of electron bunching predicts that the n th harmonic of the alternating component of the current is proportional to $J_n(nx)$; x is the bunching parameter, given by the relation $x = V_1 s \pi f / V_0 v_0$ where v_0 is the velocity of the electrons accelerated to a potential V_0 , and V_1 is the alternating voltage of frequency f . With the local oscillator tuned to a wavelength of 10 cm., maximum power in the 14th harmonic (of the order of 10^{-4} watts) was obtained with values of $V_1 = 250$ – 270 v. at $V_0 = 2000$ v., and $V_1 = 140$ v. at $V_0 = 1350$ v.; both of these results correspond to $x = 1.3$, $nx = 16.8$. The first maximum of the 14th Bessel function occurs at $nx = 16$, so that the agreement with theory is satisfactory. The theoretical pulse, the shape of which it is hoped to verify with viewing equipment, is shown in the inset to figure 1 to be of the order of 2×10^{-10} seconds long. Responses have been plotted for varying values of nx , and lie on a curve of roughly the same form as that of the Bessel function, although observation of maxima beyond the first has not yet been attempted.

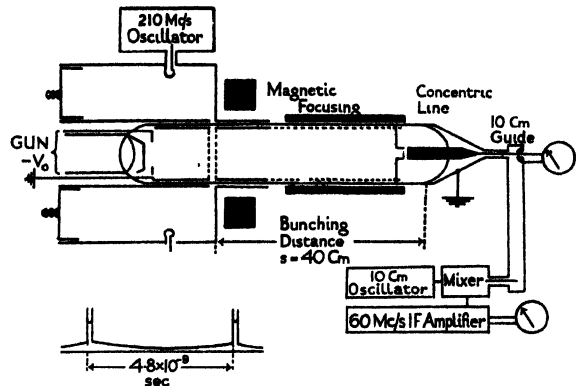
From the power observed in the 14th harmonic, and the agreement with theory, we should expect that this device will be capable of supplying pulses of variable width and of duration short enough for testing wide-band coincidence counting circuits.

The Clarendon Laboratory,
Oxford.

J. B. HASTED.

11 December 1947.

COLLIE, C. H., HASTED, J. B., and RITSON, D. M., 1948, *Phys. Soc. Proc.*, **60**, 71.



Bunching pulse generator and associated circuits in block diagram form.

Inset: Theoretical pulse

REVIEWS OF BOOKS

Radar System Engineering, edited by LOUIS N. RIDENOUR. Pp. xviii + 748: with 646 figures. (New York: McGraw Hill Book Co. Inc., 1947.) 37s. 6d. net.

During World War II a tremendous research and development effort went into the development of radar and related techniques, and resulted not only in hundreds of radar sets being available for military use but also in the collection of much valuable data and new techniques in the electronics and high-frequency fields. This effort was a splendid example of co-operative research under war-time conditions, for each new advance was rarely the work of an isolated worker, but generally the combined efforts of a very large group of people: British, American, Army, Navy, R.A.F., civilian, Government, and industry, working in different laboratories, and often at widely separated locations. A very large and important part of this combined effort was undertaken by the Radiation Laboratory of the Massachusetts Institute of Technology, and they have now played a further important part by undertaking the heavy responsibility of preparing for publication the results of this collective work. As originally planned the series comprised twenty-three volumes dealing with theory and techniques, but later it was decided to extend the series by the addition of a further five books dealing with radar and applied systems.

The present book, which is the first of the series to be published, is also the first of the five dealing with radar and applied systems. It is intended to serve as a general treatise and reference book on the design of radar systems: it deals, in fact, with the principles and practice of radar. Much of the work has been written by Professor Ridenour himself, but he has been assisted in this monumental task by many other writers, thirty-two of whom have contributed to this first volume.

The aim of the book, as stated in the preface, is to cover the collective efforts of the many workers, both in the United States and in Britain, Canada, Australia and other British Dominions, but in spite of this, the book tends to be rather overweighted with American applications of the new techniques, and this may detract from its value to the British reader. Furthermore, although various references to the original literature are given throughout the book, the references mentioned are usually to obscure reports (e.g. M.I.T. Radiation Laboratory Reports) not generally available in this country. The book would have been of more value to the British reader if reference could have been made to the excellent literature on the subject published last year by our Institution of Electrical Engineers. However, these are small points, and it can be said at once that this is a very good book, which should be of real value to all those wanting to make themselves conversant with the principles and practice of the subject.

Chapter I is a general introduction and explains clearly and concisely "what radar does" and "how it works". It then goes on to give details of the components of a radar system, the types of radar developed during World War II and the performance obtained. The author departs a little from the usual custom by referring at the outset in his explanation to a microwave pulse radar system, rather than to the simpler and historically earlier metric-wave systems. This insistence on the importance of microwaves from the start is quite reasonable. Considerably more work has gone into pulse radar than into any other kind, and by far the greater part of this work has been concerned with microwave pulse radar. Moreover, it is now recognized that for almost all radar purposes microwaves offer advantages. This first chapter is very well written and the explanations are good, but one small mis-statement has been made. On page 11, in referring to super-refraction conditions, the author states "that such phenomena are relatively rare and essentially unpredictable". This statement would be true in referring to this country; it is probably also true in referring to conditions in North America, but it is obviously not true in relation to conditions in more tropical countries. The Royal Air Force have observed super-refraction conditions repeated for several weeks in the year, for several years in succession, in certain parts of India, and so for certain parts of the world, at least, the increase in range of a radar set beyond the optical horizon is a predictable phenomenon.

Chapter II contains a treatment of the radar equation according to which the range (R) of a radar set is a function of the transmitted power (P), the gain of the antenna (G), the wavelength (λ), the radar cross-section of the target (σ) and the minimum detectable received signal (S). The author writes down his formula as

$$S = \left(\frac{PG}{4\pi R^2} \right) \left(\frac{\sigma}{4\pi R^2} \right) \left(\frac{G\lambda^2}{4\pi} \right), \quad \dots\dots (A)$$

and proceeds to discuss in a very thorough fashion the consequences of the equation, and the choice of the various parameters for particular purposes. He discusses beams of special shapes, the limitations of the receiver, receiver noise and the statistical problem, and the modification of the range equation called for when "free-space" propagation conditions are not applicable. He discusses one case in detail, namely the calculation of the maximum radar range of an aerial target when the propagating path is over a flat, perfectly reflecting surface, and shows that in this case the "free-space" formula becomes

$$S = P \frac{G^2 \lambda^2 \sigma}{(4\pi)^3 R^4} \cdot 16 \sin^4 \left(\frac{2\pi h_1 h_2}{\lambda R} \right), \quad \dots\dots (B)$$

and he discusses, in detail, how accurately the flat, perfectly reflecting conditions laid down in his analysis are fulfilled in practice, and therefore how reasonable it is to use the new equation. The treatment, so far as aerial targets are concerned, is thorough and excellent, but the author fails to point out the limitation of his analysis to ship targets, and except for a casual reference to figure 2.12, the reader is left with the impression that equation (B) may be used for ship targets. This is very unsatisfactory. The reviewer would have liked to have seen the important case of the ship target considered in detail with proper comment on the special conditions which arise in this case (namely that the target is not placed in a constant incident field, but in one which varies across the surface of the target), and the modifications to the concept of target cross-section made necessary. A satisfactory treatment of this problem was published recently by Wilkes and Ramsay (*Proc. Camb. Phil. Soc.*, 1946, p. 220).

Chapter III gives a detailed treatment of the properties of radar targets, and after a consideration of targets of simple shapes, e.g. sphere, corner reflector etc., some discussion is included on methods of target shaping and the use of absorbent materials to diminish cross-section. The chapter then goes on to discuss the radar returns from more complicated targets such as aircraft, "window", rain, land, and the sea, and this is illustrated with many excellent photographs of the radar returns actually obtained. The use of radar "land-painting" is considered very fully in connection with the important applications of this technique to navigation problems.

Chapter IV discusses the limitations of pulse radar, and Chapter V gives a detailed treatment of the many c.w. radar systems, which includes many critical inter-comparisons of the performance of c.w. and pulse systems. The c.w. systems considered include simple Doppler systems, range-measuring Doppler systems, frequency-modulation range measuring systems, multiple-target frequency-modulation systems, pulse-modulated Doppler systems etc. One small mis-statement occurs sufficiently frequently to be annoying, namely that a frequency-modulation system will work down to zero range. There is a minimum range for a F.M. system but it is not zero, but equal to $c\delta f/2F\Delta f$, where δf is the beat frequency, F is the modulation frequency and Δf is the frequency swing. This reduces to $c/2\Delta f$ when $\delta f = F$. The reviewer would have liked to have seen this point dealt with properly. The corresponding figure for a pulse system is $\frac{1}{2}c\tau$, where τ is the pulse length. For a normal F.M. system the limitation is usually of the order of 1 metre, whereas for a pulse system the limitation in range might be as great as 100 metres; hence all the author's statements about the advantages of c.w. systems for the accurate measurement of very short ranges are obviously true.

Chapter VI gives details of the presentation of radar information, and most readers will find this a valuable introduction to the various applications of radar methods. Chapter VII considers the employment of radar data, and gives examples of radar organization, viz. radar in the R.A.F. Fighter Command and the U.S. Tactical Air Commands. It is very pleasing to the reviewer to find this important material included. The technical design of a radar set is relatively straightforward to the specialist at least, but the employment of radar data in the

best possible way is far less well understood, and Professor Ridenour is quite correct when he states that the major improvements to be looked for in the use of radar during the next few years will lie, for the most part, not in technical radar design, but in the field of fitting the entire radar system, including its operational organization, to the detailed needs of the use and the user. This obviously applies to both military and peace-time applications.

Chapter VIII continues with a consideration of radar beacons, and gives a good introduction to the subject. The whole of volume 3 of the series is also devoted to this subject. Chapters IX to XIV take up the leading design considerations for the various important components that make up a radar set, and deal respectively with aerials, magnetrons, R.F. components, radar receivers, indicators and power supplies. The treatment is thorough and written throughout by specialists in their subjects. The reviewer would have liked the needs of the British reader to have been considered, by finding a few more references to British practice, and he would have expected the 70-page chapter on the magnetron developments to have mentioned that the all-important cavity resonator magnetron was invented by Professor Randall, our recent Duddell medallist, but bearing in mind the emphasis on co-operative efforts, perhaps this is not so unfair. Chapter XV is one of the most valuable in the book and gives some examples of radar system design practice. The authors (R. G. Heib and R. L. Sinsheimer) consider the complete design problem from a consideration of the formulation of requirements, to the choice of the various radar parameters, and to the detailed problems of the design of the various component parts of the equipment. This chapter has been carefully written and will repay careful study. Finally the book closes with a consideration of two new and important ancillary techniques—moving target indication and the transmission of radar displays to a remote indication by radio means.

The book is well produced and illustrated and can be recommended to any serious student of radio. It is hoped that some of the criticisms made above will be dealt with in future editions.

DFNIS TAYLOR.

Radioactivity and Nuclear Physics, by J. M. CORK. Pp. x+313. (New York : Van Nostrand Co. Inc., 1947.)

This book will be easily read—and with a fair measure of assent—by anyone familiar with the subject, yet it is a tantalizing and a disappointing book. It lacks logical form : there is a great deal that passes for information, very little which seeks to be educative. The serious student will not find that the book answers his questions, though it will certainly give rise to many, in a bewildering sequence. All the fashionable terms are freely used—curium is mentioned on p. 3, fission on p. 6, the relativistic change of mass with velocity on p. 14, and the meson and the neutrino on the following page. But that surely is the wrong place for them, unless a very casual acquaintance with words is all that the reader requires.

“It has been the aim of this text to offer an authentic historical development of each topic”: thus the preface. This much must be granted, that the aim is evident throughout; nor is its success negligible (most obvious perhaps in Chapter 3 which describes the evolution of large-scale apparatus for particle acceleration), but the historian is at fault, sometimes, and at others is uncritical. It would seem strange that the earliest reference to Rutherford's work in radioactivity is of date 1902, and the almost literal statement that the experiment of Ellis and Wooster (1927) was planned in order to test the neutrino hypothesis is a peculiar reversal of history. Yet the reader should be grateful that the historical approach is followed consistently enough, whenever occasion offers, and the copious references which appear as footnotes to the text are a good feature of the book.

At the end appendices give tables of isotopes (stable and unstable), of fundamental constants, of numerical data concerning electrons, protons and α -particles over a considerable range of energy, and of the characteristic x-ray excitation energies of the elements titanium to uranium. These tables also are good (incidentally the table of isotopes includes some data regarding recently discovered isotopes of uranium and the transuranics which, so far as the present reviewer is aware, have not hitherto been published either in this country or in America). But why exclude rhenium and bismuth from the table of x-ray excitation energies? This is rather a frivolous question, but it should set the user of the tables on his guard—for there are, perhaps inevitably, some errors as well as omissions.

From what has been said in praise of the industry which has gone to the compilation of lists of references and useful tables of data it will be evident that it is the greater pity that

the general level of the text is not better. But criticism cannot be avoided: the text too frequently appeals to the memory of the student, not to his understanding. And even his memory is not encouraged by a repetitiveness which so often introduces inconsistency: the maximum energy of the β -particles of radium E is given as 1050 kev. on p. 124 and as 1.17 mev. on p. 133, the discovery of the positron is attributed to Anderson in 1933 on p. 153—and in almost identical language the date is given as 1932 on p. 124 (incidentally, on the former page the relativistic theory of the electron is dated 1934—six years too late), the energy release in the reaction ${}^6_3\text{B} (n, \alpha) {}^3_3\text{Li}$ is quoted as 2.5 mev. on p. 177 and as 3.0 mev. on p. 181, that for the reaction ${}^3_1\text{H} (d, n) {}^3_2\text{He}$ as 3.17 mev. on p. 178 and 3.16 mev. on p. 198 (the "best" value is in fact 3.23 mev.), that for ${}^9_4\text{Be} (d, n) {}^{10}_5\text{B}$ as 4.1 mev. on p. 178 and 3.74 mev. on p. 198, and on p. 210 it is stated that the specific ionization due to an electron is independent of its energy when the energy is high, whereas on p. 229 we have the correct statement that the ionization increases with increasing energy in this range. These are pure inconsistencies, which diminish confidence when they are detected; of the outright errors, which the unwary student might not find it so easy to detect, it will suffice to quote one in full (p. 125)—"The upper limit of the spectrum for the ${}^{64}\text{Cu}$ positrons is 0.66 mev. compared with 0.58 mev. for the negatives. The difference may be due to the action of the positive nucleus on the departing particles, being an attraction for the negatron and a repulsion of the positron". It must be admitted that there is a superficial air of verisimilitude in this wholly misleading statement.

N. FEATHER.

Tables of Spherical Bessel Functions. Vol. II. Prepared by the Mathematical Tables Project of the National Bureau of Standards. Pp. xx+328. (New York: Columbia University Press, 1947.) \$7.50.

The tabulation of the functions $\sqrt{(\pi/2x)} \cdot J(x)$ is continued*, and this volume covers the ranges

$$\begin{aligned} \pm 2\nu - 29 \quad (2) \quad 43; & \quad x = 0 \quad (0.01) \quad 10 \quad (0.1) \quad 25; \\ \pm 2\nu - 45 \quad (2) \quad 61; & \quad x = 10 \quad (0.1) \quad 25. \end{aligned}$$

Apart from a few entries near the zeros, the values are given to eight or more significant figures for $x < 10$, and to seven for $x > 10$. Second differences, modified second differences, or second and fourth differences, are given to facilitate interpolation, except in ranges from $x = 0$ where the functions vary very rapidly. To enable interpolation in these latter ranges (and also in the ν direction) to be performed, values of

$$A_\nu(x) - 2^{1+\frac{1}{2}} \pi^{-\frac{1}{2}} \Gamma(\nu+1) J_\nu(x)$$

are given for

$$\begin{aligned} 2\nu - 1 \quad (1) \quad 41 \quad (2) \quad 61; & \quad \nu = 0 \quad (0.1) \quad 10; \\ 2\nu - 1 \quad (2) \quad 61; & \quad x = 10 \quad (0.1) \quad 25; \\ -2\nu - 29 \quad (2) \quad 33; & \quad x = 0 \quad (0.1) \quad 9.5 \quad (0.05) \quad 10 \quad (0.1) \quad 25; \\ 2\nu - 35 \quad (2) \quad 61; & \quad x = 0 \quad (0.1) \quad 25; \end{aligned}$$

the first group to nine decimal places, the second to (mostly) seven significant figures.

Finally the zeros of $J_\nu(x)$ and $J'_\nu(x)$ obtained by inverse interpolation from the main tables of this and the preceding volume, and the value of each function at the zero of the other, are given to six (occasionally more) decimal places.

The functions here tabulated are important in, for instance, solutions of the wave equation in spherical polar coordinates, and are to be welcomed as extending the range of the tables already given in volume I.

W. G. B.

Table of the Bessel Functions $J_0(x)$ and $J_1(x)$ for Complex Arguments. Second edition. Prepared by the Mathematical Tables Project of the National Bureau of Standards. Pp. xlv+403. (New York: Columbia University Press, 1947.) \$5.

This is substantially a reprint of the first edition, but some errors of lettering in the diagrams have been corrected and explicit relations between the functions tabulated along the 45° ray and the *ber* and *bei* functions are now included, with other minor revisions, in the introduction.

W. G. B.

* cf. *Proc. Phys. Soc.*, 1947, 59, 259.

Time and Thermodynamics, by A. R. UBBELOHDE. Pp. vi+110. (Oxford: University Press, 1947.) 6s. net.

The main theme of this well-produced little book is that, in the metrical study of the external world, vision (i.e. the use of any of the so-called ethereal vibrations) is of prime importance, and since vision depends on the absence of radiation equilibrium (in a constant temperature enclosure the world would appear as an undifferentiated uniformity) the conditions of such absence (e.g. the continual increase of entropy) are ultimately the conditions of obtaining metrical knowledge of the universe. "And since this is so it may be anticipated that the trend of events, or direction in time, which is indicated by thermodynamics has its foundations primarily in the thermodynamics of radiation" (p. 77).

The book is disappointing, both in its main thesis and in the development of it. There are other conditions, besides absence of radiation equilibrium, of obtaining metrical knowledge; for example, the temperature of the body must not exceed, say, 110° F., the atmosphere must not be deprived of oxygen, and so on. It is not clear why, among these conditions, the absence of radiation equilibrium is "prime". When we come to time (which, by the way, is very much the junior partner, notwithstanding the order of names in the title), the book is even less satisfactory. It appears to be held that entropy enables us to measure spontaneous change. "Historically, the way in which the concept of entropy first aroused an interest outside its immediate technical applications was by providing a quantitative measure for the trend of spontaneous happenings in the particular instance of the direction of spontaneous heat flow from one body to another" (p. 78). This, of course, is nonsense. The second law of thermodynamics merely requires that in an irreversible change from one equilibrium state to another the entropy must increase. It says nothing at all about the *rate* of increase, and therefore provides no quantitative measure of the trend.

Still more unexpectedly, the book abounds in gross errors. Perhaps the most surprising is in the discussion (pp. 83-6) of the expansion of a gas from a volume v_1 to a volume v_2 . It is stated that "the ratio of the probability W_1 that the molecules will all remain in the volume v_1 to the probability W_2 of their occupying uniformly the larger volume v_2 which is made available to them, is given by $W_2/W_1 = v_2/v_1$ ", and even the recital of an example of patrons choosing theatre seats, in which the absurdity is equally obvious, has not opened Professor Ubbelohde's eyes to it. Again, "the unfolding of the solar system on Laplace's dynamical theory would be completely reversed by reversing the momenta of all the masses taking part" (p. 79). Does Professor Ubbelohde really think that if the bodies revolved the other way the planets would return to a Sun which would expand back to the original nebula? Again, infra-red solar radiation "is largely responsible for the sun's heat, and for the comfortable use of glasshouses" (p. 61). Professor Ubbelohde surely knows how small a proportion of the sun's heat comes from infra-red radiation, and how negligible its part is in the greenhouse effect. Once more, the uncertainty principle is stated as $\Delta E \cdot \Delta t \geq h$, "where h is Planck's constant in radiation theory, and ΔE is the minimum uncertainty in E , Δt the minimum uncertainty in t " (p. 75). These increments are, of course, any uncertainties at all. The least value for each is zero and the greatest infinity, and only the product has a minimum value. There are other faults of the same type, of which we need add only the inexcusable transformation by which probability conceptions, first properly introduced (p. 83) as terms in which the laws of thermodynamics can be "reformulated", automatically become "the statistical foundations of the Second Law of Thermodynamics" (pp. 87, 97, 101; reviewer's italics).

It is difficult to persuade oneself that this book has really come from so competent and well-informed an author as Professor Ubbelohde, but one must believe the Oxford University Press. He is plainly guilty of contempt of subject. Fortunately a reviewer is in the position of jury and not of judge, and we gladly refrain from passing sentence.

HERBERT DINGLE.

Fundamentals of Photography with Laboratory Experiments, by PAUL E. BOUCHER. 2nd edition. Pp. xii+394+liii. (New York: D. Van Nostrand Co. Inc., 1947.)

The first edition of this work was published in 1940; the present edition has been broadened, so the preface tells us, by the inclusion of more material on colour photography

and by chapters on ultra-violet and infra-red, x-ray, and motion-picture photography. The book is intended to supply lecture and laboratory notes for a one-semester course in photography for college students. The major part consists of some nineteen chapters ranging from the historical development of photography to cinematography. There follow a series of twenty-five laboratory exercises which follow the same order as the chapters.

The book appears far too ambitious. It contains too much and as a result in the treatment of specific matters it is superficial, and many careless statements occur. The sketch of the historical development of photography is very unsatisfactory. Fox Talbot is left out entirely and Herschel is not mentioned. Niepce is mentioned, but his really important early work is overlooked. In the next chapter is a heading—"The Reduction of the Latent Image". One knows something of the development of the latent image, but we do not usually talk about its reduction. Very many more examples could be cited of careless statements, and it has to be remembered that this work is intended as a college text-book. It does not deal satisfactorily with the fundamentals of photography, while at the same time dealing very superficially with many other matters. It is surely unnecessary to provide a description and a wiring diagram of an x-ray generator when no mention is made of the use of intensifying screens in x-ray work. There is no mention of stereoscopic photography, but quite a lot about certain aspects of ultra-violet and infra-red that might well have been left out.

The general plan of the book is good and most of the information sound and well arranged; it is a great pity that it should be marred by such imperfections as are detailed above.

H. W. GREENWOOD.

Atomic Energy, by R. R. NIMMO. Pp. 201. (London: The Pilot Press Ltd., 1947.) 9s. 6d.

"The average citizen cannot be expected to understand clearly how an atomic bomb . . . works" says Professor Smyth in the preface to the official United States booklet on Atomic Energy. On the other hand, the Pilot Press and their editor, Dr. Lovell, believe that the "non-specialist public" is capable of keeping abreast of such topics as atomic energy, and they publish this book by Dr. Nimmo to prove their point. We shall not attempt an opinion on this controversial question here, but simply examine the content of Dr. Nimmo's book. It contains an excellent account of nuclear fission, chain reactions and uranium piles, leading up to the atomic bomb and atomic energy plants. A fairly conservative attitude is indicated towards the future of atomic power plants. Ten thousand tons, or more, of uranium would be needed yearly to provide the power requirements of Great Britain, and there is little up-to-date information available, apparently, regarding the size of uranium deposits. Generally speaking, all the information on atomic energy which has been released, and all the deductions which a nuclear physicist can make from the official statements, are included in the book. It can be recommended to any scientist as a competent survey, carefully written by an expert in this field. The Pilot Press turn out a pleasing volume and are welcomed into this sphere of publication. Their production would have been improved, however, if they had conformed to the standard practice of printing mathematical symbols in italics.

W. COCHRANE.

The Evolution of Modern Physics, by C. T. CHASE. Pp. ix+203. (New York: D. van Nostrand Co. Inc., 1947.) 14s.

There is always much pleasure to be derived from accounts of the many striking incidents and events in the history of physics. Galileo's work on falling bodies; Fresnel's triumph over the academic doubters; Röntgen's discovery of x-rays; Becquerel's discovery of radioactivity—these, and many others, never fail to delight. If the writer manages, at the same time, to show the use of the scientific method on these occasions, it will benefit the general reader, who is patently aware of the achievements of science, but less familiar with the discipline of thought which science requires.

Professor Chase's book contains a well-written and connected account of the history of physics, with a reasonable emphasis on the methods and thought involved, and with practically no mathematics. Some points dealing with the atom and nucleus are rather concentrated and one looks in vain for the apt analogy that would help over an awkward fence.

There are no diagrams or sketches in the book but there are a dozen plates, mainly of complicated pieces of apparatus not referred to in the text. The book is, however, a useful one, since it maintains a high standard. It will help the reputation of physics and will lead people to an understanding of the physicist's work.

W. COCHRANE.

The Electron Microscope, by V. E. COSSLETT. Pp. viii + 128. (London: Sigma Books Ltd., 1947.) 7s. 6d.

The electron microscope is hardly more than 15 years old, yet half a dozen books have been written on it. These books range from the exhaustive monograph to the short introduction for scientists who have no intention to specialize in the subject. The author of the present book now brings principles and uses of the electron microscope "before that wide and increasing public which has a lively interest in the advance of science".

The content of the book may be divided into three parts of roughly 40 pages each: (1) General introduction, (2) Principles and designs of electron microscopes, (3) Present use and future possibilities of electron microscopes.

The treatment starts in a very general way with some principles of vision, with refraction, prisms and lenses; then it proceeds to the eye, to telescopes and microscopes. In rather quick succession follow explanations on the electron, its refraction and deflection, and on general features of electron lenses. Two excellent chapters on electron microscopes contain details of magnetic and electrostatic types, of shadow- and of scanning-microscopes. Then follows a brief survey of electron microscopical methods and preparation techniques for the investigation of powders, fibres, bacteria, viruses etc. Sectioning and straining methods are discussed and replica methods for the investigation of surfaces are described. A concluding chapter contains information on limitations and possibilities of the electron microscope. These points are of great interest, but perhaps the author has here exceeded the frame of a booklet which is intended for a generally interested public.

The little volume is well illustrated by 40 figures in the text and 12 plates containing a selection of beautiful electron microscopical photographs. The name of the author—who has also written a standard monograph on Electron Optics and who is now in charge of electron microscopic developments in the Cavendish Laboratory in Cambridge—vouches for a masterly treatment of the subject. The book should be warmly recommended to everybody who is interested in electron microscopy.

O. KLEMPERER.

A University Text-Book of Physics, Vol. I: Properties of Matter, by POYNTING and THOMSON and TODD. Pp. 278. (London: Charles Griffin and Co., Ltd., 1947.) 20s.

Of the famous text-book by Poynting and Thomson, the most famous volume was the first, on properties of matter. It is appropriate, therefore, that the first volume to appear of the revision now being undertaken should be that on properties of matter. The original work is so well planned, and so good in its treatment, that one feels reluctant to see too much revision; if an author desires to treat the whole subject differently, it is better that he should write a completely new book. Fortunately, Prof. Todd, the reviser of this volume, and an old student of *both* senior authors, has treated the original text with due respect, and has revised only to the extent that was called for by the developments in the interval since first publication. His chief additions are in the chapters on gravitation, surface tension, and viscosity, and he has certainly made the book a suitable introduction for present-day students. Like others of Griffin's post-war publications, the book is printed on thinner paper than they used to adopt, and the result is to make a pleasanter book to handle.

J. H. A.

Frontiers of Astronomy, by DAVID T. EVANS. Pp. 175. (London: Sigma Introductions to Science, 1946.) 6s. net.

This is a very good little sketch of the present state of astronomy, giving concise information concerning its chief problems, methods and results. The book is written in a racy, conversational style and is illustrated by a number of diagrams in keeping therewith. A remarkable amount of knowledge is compressed into a small space without introducing obscurity, and although there are, unfortunately, a rather large number of slips, they are mostly of slight importance, and the reader can be satisfied that the impression he gets is an essentially true one. As present prices go the book is strikingly good value for the money.

H. D.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 60, PART 5

1 May 1948

No. 341

A Quantitative Study of the Expansion Method for Liquefying Helium

By G. L. PICKARD * AND F. E. SIMON

Clarendon Laboratory, Oxford

* Now at University of British Columbia, Vancouver

MS. received 12 December 1947

ABSTRACT. The yield of the Expansion Method for producing liquid helium has been measured for various starting conditions and the results are presented in the form of two diagrams.

In addition some data are given on the volumes and specific heats of compressed helium gas; although these are only of a preliminary nature, they give all the data necessary for the design of such an expansion liquefier.

The entropy diagram has been extended to higher pressures and the actual yield compared with the "ideal" yield of a piston and cylinder arrangement. It appears that the simple expansion method results in an efficiency which is 60% of the ideal when starting from the usual temperature and pressure conditions. The reasons for this high efficiency are discussed.

§ 1. INTRODUCTION

THE principle of the Expansion Method (Simon 1932a, b, 1935, 1936, 1937, Simon and Ahlberg 1933) of liquefying gases is a very simple one. The gas is compressed isothermally into a container C (figure 1) called the expansion chamber, which is pre-cooled in a low temperature bath A to the lowest temperature possible. The container is then isolated by evacuating B and the gas expanded slowly through valve V outside the apparatus. After this expansion a certain proportion of the gas remains as liquid inside the container. Such a method would be quite impossible with most gases, and is only feasible for gases with very low boiling points. The reason is that at ordinary temperatures, say room temperature, the container has always a much higher heat capacity than the gas which it contains. With falling temperature, however, conditions change radically because, first, at constant pressure a given vessel contains more gas at low temperatures, and secondly, the specific heat of solid bodies disappears with the approach to absolute zero.

Table 1 gives the heat capacities of a steel container of 150 cm³ capacity, designed (with the conventional safety factor) to stand up to 100 atmospheres, together with the heat capacities of the helium which it would contain at two temperatures.

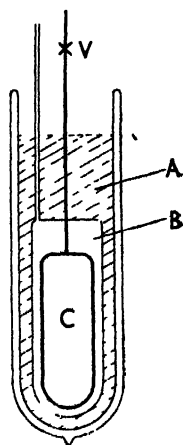


Figure 1.

Table 1. Heat capacities (cal/deg.) of steel container and helium

	Room temp.	10° K.
Steel container	50	0.1
Helium at 100 atm.	2	20
Ratio	25	0.005

Thus while at room temperature the container has 25 times the heat capacity of the gas, at 10° it has only $\frac{1}{500}$; this represents a change by a factor 5000 of the ratio between the capacities of the container of which the greater part is due to the falling off in the specific heat of the steel. It is still possible to liquefy hydrogen (Simon *et al.* 1935) by this method, but its main application is in the liquefaction of helium.

The method has been used extensively at the Clarendon Laboratory and also in a few other laboratories (for a full list of work see Simon 1940). In most of the experiments carried out, the appropriate low-temperature apparatus has been attached to the expansion chamber. This method of working has many advantages, chiefly that the whole apparatus is surrounded throughout by liquid hydrogen and thus the heat losses are kept very low. An expansion chamber of about 100 cm³ capacity provides enough helium for an experiment of 10–20 hours' duration, and as a result, expansion chambers larger than about 100 cm³ have seldom been used. The expansion method is therefore generally regarded as a small scale method, but there is in fact no limitation to the size of an expansion liquefier. Just before the war the first liquefier of this type was built in which liquid helium could be siphoned over from an expansion chamber of 450 cm³ into an exterior container (Cooke *et al.* 1939). Since the war another such liquefier has been described (Scott and Cook 1947) of a capacity of 400 cm³, and one of 1½ litre capacity is now under construction at the Clarendon Laboratory. With the widening interest in this method we think it worth while to publish some data, obtained in 1936, from which it is possible to decide upon the best starting conditions. Some earlier experiments on the liquid yields have been described by Simon and Ahlberg (1933) and a preliminary account of the present results was given by Simon in 1936.

The main object of these experiments was to determine the effect of varying starting conditions on the yield of liquid helium, and while doing so, to obtain as much data as possible about the diagram of state and thermal properties of helium, of which little was known at the time in the region of operation of the expansion method. It should be emphasized, however, that these latter experiments have been regarded only as auxiliary and were not carried out with high accuracy.

§ 2. GENERAL ARRANGEMENT

For the experiments an expansion chamber (enclosed in a vacuum case) was fitted with an electrical resistance thermometer and heating coil and the helium was allowed to expand into a large reservoir of known volume. Measurements of the pressure and temperature of the gas in the expansion chamber and of the gas expanded were made at intervals during the expansion. The volume of helium remaining as liquid after expansion to the normal boiling point was measured by heating the cylinder and measuring the volume of gas obtained. A series of seven such experiments from different starting conditions gave sufficient data for the construction of curves showing the relations between

the starting conditions and the yield of liquid obtained. The heating was carried out electrically so as to measure the heat capacity of the helium gas up to about 20° K. before it was let out into the reservoir for measurement.

§ 3. DETAILS OF THE APPARATUS

The expansion chamber was the hemispherically ended cylinder of Hecla/AMF steel used in the previous work on hydrogen (Simon *et al.* 1935); it had an internal volume of 144 cm³ and weighed only 212 gm. The heater and thermometer wires were wound directly on to the cylinder and attached by bakelite varnish, which was not subsequently hardened by baking. The cylinder was then covered with bright aluminium foil to reduce radiation effects. The thermometer coil of No. 47 s.w.g. Eureka wire was of about 500 Ω resistance and the heater coil of No. 40 s.w.g. Eureka wire was of about 200 Ω resistance. The wires were wound in longitudinal zig-zags up and down the side of the cylinder to reduce the risk of their being strained by distension of the cylinder at high pressures.

The single tube to the cylinder was of cupro-nickel and served to support the cylinder within a brass vacuum case, which was immersed in a bath of liquid hydrogen. The temperature of the bath could be reduced to the triple point (14° K.) by reducing the vapour pressure over the liquid. All the necessary connecting tubes for the helium supply etc. were of as small a bore as practicable in order to reduce the dead volume correction. The pressure of the helium in the cylinder was measured by means of a small volume bent-tube manometer and the pressure in the reservoir by a precision bent-tube manometer calibrated against a mercury U-tube manometer.

The resistance thermometer was calibrated from 10°–20° K. using the known vapour pressure–temperature relation for hydrogen after liquefying hydrogen in the cylinder; from 2°–5° K. it was calibrated against the vapour pressure of liquid helium, and in the intermediate region from 5°–10° K. by using the cylinder as the bulb of a helium gas thermometer.

In performing the experiments, the vacuum case, precooled with liquid air, was immersed in a bath of liquid hydrogen and a small quantity of helium gas admitted to permit thermal exchange between the cylinder and the bath. Helium was compressed into the expansion chamber to a pressure of about 160 atmospheres and the temperature of the bath lowered to the desired starting temperature. When thermal equilibrium had been attained the exchange gas was pumped from the vacuum case to isolate the cylinder. The heat inflow along the connecting tube could be neglected during the expansion.

The expansion was effected in stages, the gas being permitted to expand into the reservoir and the pressure and temperature of the gas in the cylinder and in the reservoir measured at each stage. On reaching a pressure of 1 atm. in the cylinder the valve to the reservoir was closed. The specific heat of the helium remaining in the cylinder was measured up to 20° K. and the volume of this helium determined by releasing it into the reservoir. From this volume was calculated how much had been present as liquid at the boiling point after correction for the helium in the form of vapour above the liquid.

The "yield" of liquid was then defined for practical purposes as the percentage of the volume of the cylinder which remained filled with liquid after the expansion. The range of starting conditions covered during the series of expansions enabled

a yield diagram to be constructed showing the relationship between starting conditions and the yield obtained. Furthermore, from the measurements of corresponding pressure, temperature and amount of gas contained in the known volume of the cylinder it was possible to draw a diagram of state for the gas under the conditions of the expansion method.

With the above apparatus it was practicable to use pressures up to 160 atm. and temperatures as low as 14°K ., which resulted in yields up to about 45%. It was found impossible to reach lower starting temperatures by pumping the bath alone, and therefore in order to extend the measurements a second apparatus was used. In this apparatus the helium cylinder had a hydrogen chamber attached to it and by reducing the pressure over this hydrogen the temperature of the helium before expansion could be reduced to 11° or 10°K . With this apparatus it was possible to obtain yields up to 80%. The measurements with the second apparatus were rather less accurate than those with the first and no attempt was made to obtain any diagram of state or specific heat results. The effect on the yield of the residual solid hydrogen in the hydrogen chamber was calculated and also checked by experiment. This factor introduced a correction of a few per cent to the yield obtained. The temperatures with this apparatus were measured by means of a simple gas thermometer of the Simon type (Mendelssohn 1931).

With both sets of apparatus the necessary corrections were made throughout for the dead volume of the tube system and a correction was applied to allow for the volumes of the cylinders for thermal contraction. The correction for distension of the cylinder due to the internal pressure was negligible except at the highest pressures.

§ 4. RESULTS

The experimental data for the pressure-temperature relationships during expansion are represented in figure 2, the curves with roman figures being those

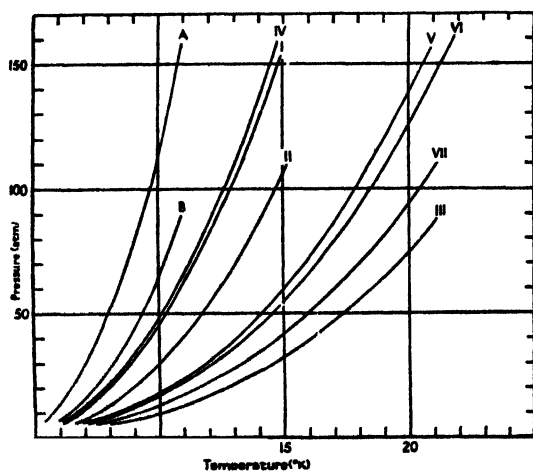


Figure 2. Experimental pressure-temperature data observed during expansions.

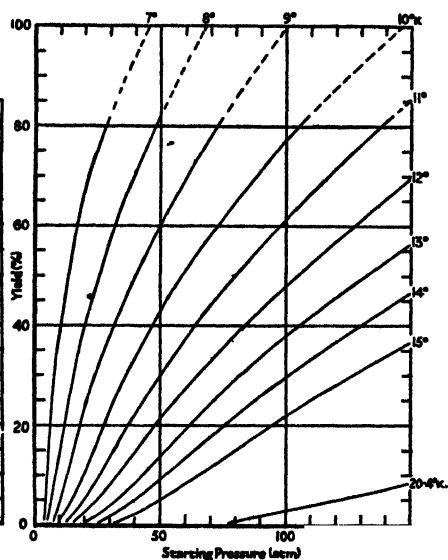


Figure 3. Yield of liquid helium by the expansion method. (Percentage of volume of expansion chamber remaining filled.)

made with the main apparatus, while A and B represent the expansions made with the second apparatus.

The most generally convenient method of plotting the *yield results* is that of figure 3 in which the yield is plotted against the starting pressure for a series of starting temperatures from 7° to 20·4°K. The extent to which the yield is increased by raising the starting pressure and lowering the starting temperature is clearly seen. A comparison between these results and those of Simon and Ahlberg (1933) was made. The curved form of the present yield isotherms is very marked, while Simon and Ahlberg whose experiments were of a less precise nature than the present ones and did not extend to such high yields, represented their yield isotherms as straight lines as a first approximation. A closer inspection of the original experimental results shows that a better representation is given by yield isotherms of similar form to the present ones. With this readjustment the agreement between the earlier results and the present ones is good. Some figures for the yields are given in table 2.

Table 2. Yield of liquid helium obtained by the expansion method expressed as the percentage of the volume of the expansion chamber remaining filled after the expansion

	Starting temperature (°K.)				
	8°	10°	15°	20·4°	
Starting pressure (atm.)	150	100	37	10	
	> 100	76	22	3	
	50	82	43	5	0

Some of the *diagram of state* data are plotted in figure 4 in the form of atomic volume isotherms at pressures up to 150 atmospheres. Values for the atomic volume are also given in table 3.

Table 3. Atomic volume of helium gas (cm³)

Temp. °K.	Pressure (atm.)						
	5	10	20	30	50	100	150
8	88·0	45·5	32·0	28·0	—	—	—
10	137·0	64·5	38·5	31·5	26·5	22·0	19·5
15	240·0	112·0	60·5	43·5	32·0	23·5	21·0
20	—	171·0	85·0	—	39·0	27·0	23·0

The figures in table 3 agree with the data of Zelmanov (1944) within the accuracy of the present measurements (estimated as $\pm 3\%$) except in the region of 20 atm. pressure and 8°–12°K. where the present figures are about 6% greater than those of Zelmanov.

The *atomic heat* of helium gas at constant volume was measured at various concentrations. These results show a considerable scatter in places and this is attributed to the varying conditions under which the readings were taken. These measurements were made at the end of the experimental "runs" and

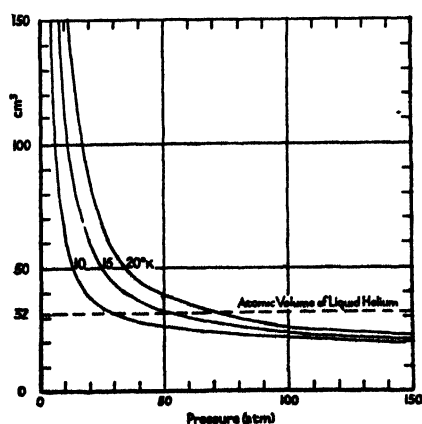


Figure 4. Atomic volume of helium gas.

very often a shortage of liquid hydrogen caused the level in the bath to fall below the upper end of the vacuum case. In consequence the temperature drift of the cylinder increased and varied, due to increased conduction along the supporting tube, and to a reduction in the efficiency of the thermal insulation caused by desorption of gas inside the case. Although the accuracy is not higher than $\pm 3\%$, we think it worth while to give in table 4 the values of the smoothed curves for two different concentrations as hardly any other measurements have yet been made in this region.

Table 4. Atomic heat at constant volume (cal/deg.)

	Temperature ($^{\circ}$ K.)				
	7'	10 $^{\circ}$	15	20'	
Density	5	3.2	3.0	3.0 ₅	3.1
(gm.atoms/litre)	14	2.7 ₆	2.8	2.9 ₅	3.0 ₅

The only direct measurements in this region which are available from other sources are those of Eucken (1916) at 18 $^{\circ}$ and 22 $^{\circ}$ K. for a concentration of 9.3 gram atoms per litre, which give a value at both temperatures of 3.0 cal/deg.

Some specific heat measurements were made below 7 $^{\circ}$ K. but these are not quoted as they are considered unreliable on account of the phenomenon of bad heat exchange through the helium in the neighbourhood of the critical point. No effect of this nature took place during the expansion because then the heat absorption was homogeneous throughout the whole volume of gas. In view of the fact that temperature equilibrium during the expansion was established in less than 20 seconds at each stage, it was considered that the thermal exchange between the gas, cylinder and wires was good and that poor thermal conductivity of the material of the cylinder could not be responsible for the effects noted above when heating the cylinder and contents after the expansion.

§ 5. DISCUSSION

Figure 3, giving the *yield* isotherms, provides all the information needed for the construction of an expansion liquefier. The data are represented in a somewhat different form in figure 5, in which "isoyields" are plotted as functions of starting pressure and temperature. The shaded areas indicate the regions of zero yield and yield above 100 $^{\circ}$ o.

An approximate formula for the (percentage) yield isotherms of figure 3 is

$$\text{Yield} = 5.5 p_{\text{atm.}}/T + 10(12 - T)$$

within the limits $p = 75$ to 150 atm. and $T = 10^{\circ}$ to 12 $^{\circ}$ K., which covers the region of greatest practical interest. This formula represents the yield isotherms to an accuracy of $\pm 2\%$ yield.

As can be seen from the diagrams, the yield increases considerably with pressure. A practical limit to the pressure is set by the dimensions of the expansion chamber and the connecting

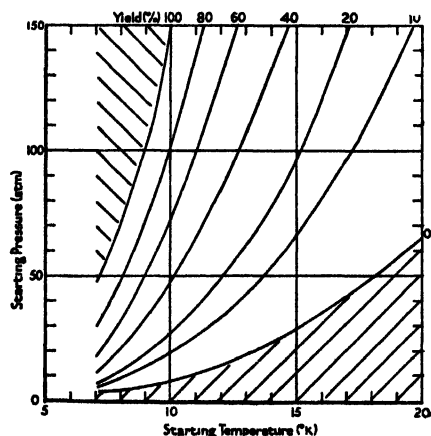


Figure 5. Yield diagram, curves of constant yield.

tube. We generally use pressures between 100 and 150 atm., and as very satisfactory yields can be obtained with these pressures at starting temperatures of the order of 10° – 12° , which are easily obtainable, there does not seem to be any point in going to higher pressures. The question whether a combination of low temperature and low pressure, or higher temperature and higher pressures, is preferable, and under what circumstances, has previously been fully discussed by Simon (1940), and we can, therefore, refer to this paper.*

The form of the *diagram of state*, figure 4, presents no unusual features. It is of great interest, however, that at fairly low pressures volumes are shown which are considerably lower than that of the liquid at the boiling point. This is due to the fact that liquid helium is "blown up" by its considerable zero-point energy (Simon 1934) to a volume about four times that which would be predicted from the data of gas kinetics. The fact that at certain starting conditions the gas has a smaller volume than that of the liquid is also the explanation for the result that we can get yields of above 100% (measured as the degree of filling of the vessel).

The *entropy diagram* of helium has been published by Keesom (1942) for a range of temperatures and pressures. An extension of this range is possible by employing the values from the isentropic pressure–temperature curves obtained during the expansions made in the present experiments (figure 5). By this means isobars for pressures up to 160 atm. have been drawn on the entropy diagram; some figures are quoted in table 5. It should be noted that the entropy is given for 1 gram atom and that the zero of entropy has been taken as that of the liquid at the boiling point (4.22°) and 1 atm.

Table 5. Entropy of helium gas (cals/gm. atom $^{\circ}$ K.)

Temp. $^{\circ}$ K.	Pressure (atm.)						
	160	140	120	100	80	60	40
7.5	—	—	—	—	–0.15*	+0.05*	0.45
10	0.20	0.30	0.40	0.50	0.75	1.10	1.65
15	1.45	1.60	1.85	2.10	2.50	3.05	—
20	2.70	2.95	3.20	3.60	4.00	—	—

* Slightly extrapolated.

The "Ideal" Yield

These extended data cannot be regarded as very accurate but are sufficient to permit calculations to be made as to the yields which would be obtained by the operation of an expansion in an ideal piston and cylinder process. For suppose that 1 gram atom of helium is compressed into a cylinder by a piston and that after an ideal adiabatic expansion to 1 atm. pressure has taken place the helium remains in the form of a fraction x as liquid and $(1-x)$ as gas at the normal boiling point. Then if the entropy of the compressed gas at the starting

* In 1939 Van Itterbeek published some data on the liquid yields of a small number of expansions which were considerably higher than in the data given by us. As one of the authors has shown (Simon 1940), there are valid reasons why Van Itterbeek's experiments might have been affected by erroneous determinations of the temperature. Later, Van Itterbeek (1943) published the results of some new determinations at conditions which resulted in a rather low yield; they agree with our values within the limit of error. It should be added that we have now made more than 1000 expansions at the Clarendon Laboratory, in most of which a rough check on the yield was taken, and that we have never found any disagreement with the results of our diagram.

conditions is S_l and those of the liquid and gas at the boiling point are S_l and S_g respectively, we have, since the expansion is an isentropic process,

$$S_l = x S_l + (1 - x) S_g.$$

Now if we put $S_l = 0$, then $S_g = 4.65$ entropy units and we have

$$x = 1 - S_l/4.65.$$

To express this as "yield" as defined previously it is necessary to compare the volumes occupied by a fraction x of 1 gm. atom of liquid with that of 1 gm. atom of gas at the starting conditions used. This latter would be the volume of the expansion chamber in a practical liquefier.

Since the atomic volume of the liquid is 32 cm^3 , the volume occupied by x gm. atoms is $32x \text{ cm}^3$ and if the volume occupied by 1 gm. atom of gas at the starting conditions is V_1 then the yield for an ideal expansion is

$$Y_{id} = 32x \cdot 100/V_1 \% = (32/V_1)(1 - S_l/4.65)100 \%.$$

From this expression it is found that the ideal yield (Y_{id}) may be greater than 100% in the case of good starting conditions for the reason already mentioned, namely that the atomic volume of the compressed gas is very considerably less than that of the liquid. In a practical apparatus of course the maximum yield is 100% as any surplus evaporates at once in the connecting tubes.

It is interesting to compare the observed yields with these ideal yields for given starting conditions. In figure 6 are plotted the values of the ideal yield as calculated above and the observed yield, for starting temperatures of 10° , 15° , 20° K . Naturally the observed yield approximates to the theoretical yield at lower starting temperatures and higher pressures. It is, however, interesting to see how high a percentage of the ideal yield can actually be obtained with such a simple arrangement, about 60% , when starting from the usual pressures and temperatures. There are two reasons: firstly, the heat capacity of the containers is practically negligible in comparison with that of the working gas—we have the advantage of working with "mathematical" walls; secondly, pressures as low as 100 atm. are very high "reduced" pressures for helium, as its internal pressure is of the order of 20 atm.; thus when working with helium we have, at 100 atm. pressure, conditions corresponding to about 10 000 atm. pressure for nitrogen, which has an internal pressure about 100 times as great.

The data already given enable us to calculate the amount of gas leaving the apparatus during an expansion, which has to be taken up by the gasometer. No diagrams have been prepared, but it should be mentioned that under average working conditions, say 100 to 150 atm. and 10 – 12° , about half the gas leaves the apparatus during the expansion.

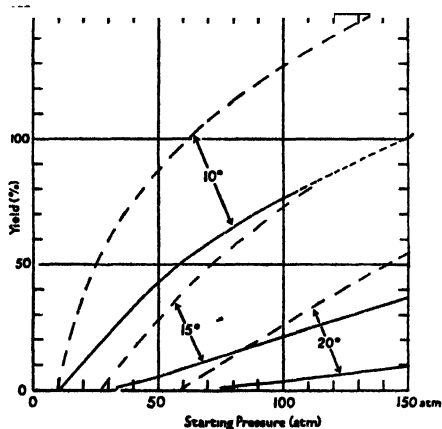


Figure 6. Comparison of ideal and observed yields from expansion method for helium liquefaction.

— — — Ideal yield of cylinder and piston.
 — — — Observed yield of expansion method.

As has been indicated in earlier work (Simon and Ahlberg 1933), an improvement of the yield can be obtained by making use of the Joule-Thomson effect. Our data on the equation of state have been used to calculate the Joule-Thomson inversion curve which is found to be in quite good agreement with the direct experiments of Zelmanov (1940). A calculation shows that in general the yield from a simple expansion is so high that the gain which would result from using the Joule-Thomson effect as well would not be worth the complication of the apparatus and technique which would be necessary.

For the larger expansion apparatus which we are now contemplating at the Clarendon Laboratory, giving a yield of one or more litres per expansion, the position may however be different, and we are studying this matter at the present moment. We shall also postpone a discussion on the efficiency of the method from the point of view of consumption of liquid hydrogen until these larger pieces of equipment are in operation.

REFERENCES

- COOKE, A. H., ROLLIN, B. V., and SIMON, F., 1939, *Rev. Sci. Instrum.*, **10**, 251.
EUCKEN, A., 1916, *Verh. dtsh. phys. Ges.*, **18**, 4.
KEESOM, W. H., 1942, *Helium* (Amsterdam), p. 247.
MENDELSSOHN, K., 1931, *Z. Phys.*, **73**, 494.
SCOTT, R. B., and COOK, J. W., 1947, *Phys. Rev.*, **72**, 161.
SIMON, F., 1932 a, *Z. ges. Kaelte-Ind.*, **39**, 89 ; 1932 b, *Phys. Z.*, **34**, 232 ; 1934, *Nature, Lond.*, **133**, 529 ; 1935, *Proc. Roy. Instn.*, **28**, 155 ; 1936, *Proc. 7th Int. Congr. Refrig.*, **1**, 367 ; 1937, *Physica*, **4**, 886 ; 1940, *Ibid.*, **7**, 502.
SIMON, F., and AHLBERG, J. E., 1933, *Z. Phys.*, **81**, 817.
SIMON, F., COOKE, A. H., and PEARSON, N., 1935, *Proc. Phys. Soc.*, **47**, 678.
VAN ITTERBEEK, A., 1939, *Physica*, **6**, 728 ; 1943, *Ibid.*, **10**, 90.
ZELMANOV, J., 1940, *J. Phys. U.S.S.R.*, **3**, 343 ; 1944, *Ibid.*, **8**, 135.

Penetration of Magnetic Field into Superconductors : I. Measurements on Thin Cylinders

BY M. DÉSIRANT AND D. SHOENBERG

Royal Society Mond Laboratory, Cambridge

MS. received 13 December 1947

ABSTRACT. Absolute values of the penetration depth λ of a magnetic field into a superconductor at different temperatures can be obtained from the relative values indicated by measurements on colloids, if $\Delta\lambda$, the difference between λ at a variable and at a fixed temperature, is known. Various methods of measuring $\Delta\lambda$ are briefly reviewed, and a detailed account is given of one particular method in which $\Delta\lambda$ is deduced from the changes of susceptibility of thin cylinders with temperature. The changes found for mercury cylinders were, within experimental error, inversely proportional to the radii as simple theory demands, and the temperature variation was consistent with that to be expected from the colloid results. The absolute value of λ at 0° K. is estimated as 7.6×10^{-4} cm. In the case of tin the values of $\Delta\lambda$ measured by the present method were in rough agreement with those of other methods, but since no colloid results were available, the absolute value of λ remains undetermined. The mercury results are discussed also in relation to critical field measurements on thin films. Some preliminary results bearing on the possible dependence of λ on magnetic field are mentioned ; it is concluded that λ does not vary appreciably in fields up to half the critical value.

§ 1. INTRODUCTION

WHEN a superconductor is in a magnetic field, currents flow on its surface in such a way as to keep the magnetic field out of the interior of the superconductor. These currents require a certain depth in which to flow, and this is also the depth to which the applied magnetic field penetrates the superconductor. If the size of the superconductor is large compared with the penetration depth λ , it behaves to a close approximation as if it had zero magnetic permeability or diamagnetic susceptibility $1/4\pi$. If, however, the size is comparable to or less than λ the superconductor has a weaker diamagnetism.

In the phenomenological theory of F. and H. London (1935) the penetration is described by the differential equation

$$\nabla^2 H = H/\lambda^2 \quad \dots\dots(1)$$

and λ is given in terms of the number, n , of superconducting electrons per unit volume, their effective mass m , and their charge e , by

$$\lambda^2 = mc^2/4\pi ne^2. \quad \dots\dots(2)$$

Although the detailed form of penetration law predicted by (1) has not yet been proved experimentally, equation (2) shows the theoretical importance of λ , in giving the ratio m/n for the superconducting electrons.

The first direct experimental evidence about λ came from measurements of the magnetic properties of small mercury particles in colloids (Shoenberg 1940) and showed that λ was of order 10^{-5} cm. and increased rapidly as the transition temperature was approached. Similar evidence came from measurements of the critical fields of thin mercury films (Appleyard *et al.* 1939), though the interpretation in this case involved certain additional assumptions about the thermodynamics of the destruction of superconductivity and was therefore less certain. The colloid results, though indicating the order of magnitude of λ , could not give a precise absolute estimate since the average particle size was not accurately known, and even if it were, an absolute estimate would involve assuming the truth of (1); they did, however, give a fairly accurate measure of the variation of $\lambda(T)/\lambda(T_0)$ with temperature T , where T_0 is a fixed temperature. If it were possible to measure also $\lambda(T) - \lambda(T_0)$, which we shall denote by $\Delta\lambda$, the combination of such measurements with those of $\lambda(T)/\lambda(T_0)$ would at once give the absolute size of λ .

Fortunately, several methods are possible for measurement of $\Delta\lambda$, and we shall indicate three which have been tried successfully :

(1) The diamagnetic susceptibility χ of a long cylinder of radius r much larger than λ , should be slightly less than the value $\chi_0 (= 1/4\pi)$ for r infinite. As is obvious from elementary considerations, and also follows from the appropriate solution of (1), we should have

$$\chi/\chi_0 = 1 - 2\lambda/r, \quad \dots\dots(3)$$

and if r is of the order 10^{-3} cm., $2\lambda/r$ should be of order 2%. In principle, accurate measurement of χ should give λ absolutely, but this is very difficult in practice, and, as will be explained later, only relative changes of χ as the temperature is varied can be measured with sufficient accuracy. If we denote $\chi(T_0) - \chi(T)$ by $\Delta\chi$, we have in fact

$$\Delta\chi/\chi \simeq \Delta\chi/\chi_0 = 2\Delta\lambda/r, \quad \dots\dots(4)$$

and so accurate measurements of $\Delta\chi/\chi$ will give $\Delta\lambda$ absolutely. This method

should work also with thin superconducting plates, but some preliminary experiments showed that the results were very sensitive to slight non-parallelism of the plate and the field, and the geometry of the measuring apparatus will have to be improved before reliable results on plates can be obtained. A full account of these experiments will be given in the present paper; a preliminary note has already been published (Desirant and Shoenberg 1947).

(2) The mutual inductance of two coils wound on a superconducting core should decrease slightly as the temperature is lowered, owing to the decrease of λ , and the change should be proportional to $\Delta\lambda$. This method was tried by Casimir (1940), but owing to a technical complication the experiment appeared to give much less temperature variation of $\Delta\lambda$ than was to be expected from the colloid results. A repetition of the experiment, avoiding the technical complication, has been made by Laurmann and Shoenberg and has confirmed the validity of the colloid results, and has also given interesting evidence for anisotropy of penetration depth in single crystals. This work will be fully described in the second paper of this series; a preliminary note has already been published (Laurmann and Shoenberg 1947).

(3) Pippard (1947) has used an R.F. method which measures effectively the difference between λ in the superconducting state, and the skin depth δ_n of ordinary eddy currents when superconductivity has been destroyed by a magnetic field. The value of δ_n is not accurately known, but for tin it is almost certainly independent of temperature, so that measurements of $\lambda - \delta_n$ at various temperatures give $\Delta\lambda$ as a function of temperature; the results (which were in fact the first for tin) agree well with those of the other two methods. For mercury, however, δ_n varies with temperature, so that the method cannot be used to give $\Delta\lambda$ until a complete theory of the skin effect at high frequencies has been developed.

Further work is in hand to extend and improve the colloid method of studying λ , and also to investigate the validity of the London's penetration law (equation (1)) and to look for a possible field dependence of λ . This will be dealt with in later papers in this series.

§2 EXPERIMENTAL DETAILS

(i) *Method of Measurement*

The method of measuring susceptibility was the same in principle as that used in the colloid experiments, but considerable improvements in sensitivity and accuracy were necessary before it was suitable for the present purpose. An account of these improvements was given at the Physical Society's conference in 1946 (Shoenberg 1947), and only those features of the method important for the present purpose will be described in detail. The principle of the method is to displace the specimen smartly in a uniform magnetic field from the centre of one coil to the centre of a similar oppositely wound coil, connected in series with the first and a ballistic galvanometer. The small change of flux linkage produced by the movement, if the specimen is magnetized, then gives a galvanometer deflection proportional to the magnetic moment of the specimen; two coils, rather than one, are used to minimize the disturbing effects of any fluctuations of current in the solenoid producing the uniform magnetizing field.

The main improvements introduced for the present experiments were (i) to immerse the coil system in liquid helium (instead of liquid nitrogen as in the

colloid experiments) thus enabling a greater number of turns to be used without increasing the circuit resistance, (ii) to use photoelectric amplification of the galvanometer deflection, thus increasing the deflection about 100 times, and (iii) to use a null method in which the magnetic moment of the specimen is compensated by the opposite magnetic moment of a small current in a 5-turn coil surrounding it. The overall sensitivity of the apparatus was about 3×10^5 mm. per c.g.s. unit of uncompensated magnetic moment, and the zero instability was of order 1 mm. in low magnetic fields and of order 1 cm. in a field of 300 gauss.

In order to understand how the measurements were actually carried out we shall now describe the null method in greater detail. As shown in figure 1, the current in the 5-turn coil is taken from across a fixed resistance a in the solenoid circuit and is therefore a fraction $a/(R+a+b)$ of the solenoid current, where R is the resistance in the decade box which controls the fraction, and b the resistance of the 5-turn coil and its leads. When the decade box is set so that the magnetic moment of the specimen is compensated, we have

$$Aai/(R+a+b) = \chi Vki, \quad (5)$$

where χ is the susceptibility of the specimen defined as I/H , V is its volume, k the field per unit current of the solenoid, i the solenoid current, and A the area turns of the 5-turn coil. Thus independently of the magnetic field-strength, the value of χ is inversely proportional to $R+a+b$. Our conditions were such that R was a few hundred ohms when the specimen was superconducting, while a was about 5 ohms and b about 1 ohm. As the liquid helium fell, the resistance of the leads and, therefore, b increased by about 0.3 ohm, and this change was allowed for where its effect was more than 0.1%.

It will be noticed that equation (5) assumes implicitly that the whole magnetic moment of the specimen is due to the solenoid field. In practice there are two sources of magnetic moment independent of i ; first the earth's vertical field is present in addition to the solenoid field, and secondly there may be small residual moments due either to ferromagnetic impurities or to frozen-in moments arising from previous magnetization cycles of the superconductor. Ferromagnetic effects may be very nearly eliminated by a suitable cycle of operations above the transition temperature of the superconductor; it was found that if a negative field of order 1000 gauss was switched on and off an appreciable negative remanence was obtained and this could then be eliminated by switching a positive field of about 300 gauss on and off. Provided no negative fields were used and positive fields did not exceed 300 gauss the remanence remained zero subsequently. Frozen-in moments were eliminated by always

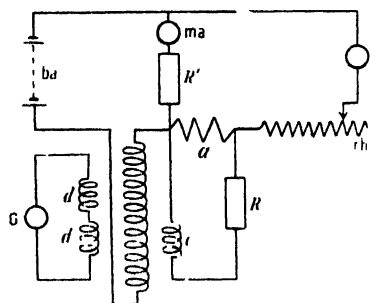


Figure 1. Schematic circuit diagram.

- ba 80 volt battery supply.
- ma milliammeter.
- am ammeter.
- G galvanometer system.
- rh rheostats controlling current in s.
- s long solenoid; 92.4 gauss amp.
- R' resistance box controlling compensating current of earth's vertical field.
- R resistance box controlling current in c.
- d d multi-turn coil in 2 opposite halves (22 000 turns in each half) located symmetrically at centre of s.
- c c 5-turn coil movable from top to bottom half of d.
- a a fixed resistance (5.2 ohm).

warming the specimen above its transition temperature before a new series of measurements was started. Finally the earth's vertical field was compensated as well as possible by passing through the solenoid an additional current which was not controlled by the main rheostats and did not pass through the resistance a or the ammeter. This compensation was achieved by making the specimen superconducting and adjusting R' (figure 1) until with no current in a the deflection was zero for pulling the specimen.

The limitation to complete elimination of the various effects just discussed was the zero instability of the galvanometer. Since it was not possible to detect reliably deflections smaller than about 1 mm., appreciable errors could still be caused where the measurement was in very low solenoid fields. Suppose that the residual uncompensated effects are such as would correspond to a magnetic moment of the specimen x (which by itself would give a deflection comparable to the zero instability) then equation (5) must be replaced by

$$Aa/(R+a+b) = \chi V k + x/i \quad \dots\dots(6)$$

and we see that such uncompensated effects can be allowed for by plotting $1/(R+a+b)$ against $1/i$ and extrapolating to zero $1/i$. This was in fact always done, and as will be seen from the examples shown in figure 2, a linear plot is indeed obtained; the value of $1/(R+a+b)$ for zero $1/i$ is proportional to the true value of χ , and the slope of the graph is a measure of λ , the uncompensated extra-neous effect. Confirmation of the validity of this procedure was obtained by deliberately increasing x (by deliberate uncompensation of the earth's field) and as can be seen from figure 2 a steeper graph is obtained but still passing through the same point on the axis of $1/(R+a+b)$. In the actual measurements we always made x as small as possible, and it is satisfactory that the linear plots were all nearly horizontal and about as often with positive as with negative slopes.

The exact setting of R to give zero galvanometer deflection proved rather difficult owing to the appearance of "double kicks" close to the balance position. These "double kicks" arise from the fact that the galvanometer has a finite period and so does not provide completely effective integration of the E.M.F.s. developed in the coil. Even when the

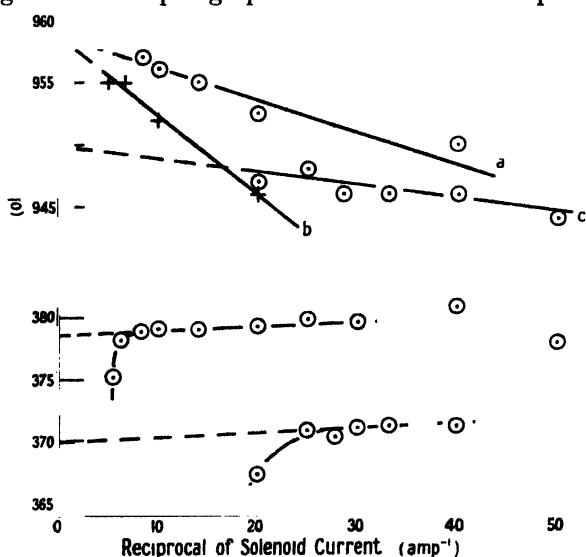


Figure 2. Illustrating extrapolation method of deducing true value of $1/(R+a+b)$.

(a) S5 : 2.16° K., (b) S5 : 2.16° K., but earth's vertical field slightly uncompensated, (c) S5 : 3.649° K., (d) H5 : 4.044° K., (e) H5 : 4.127° K.

A correction for paramagnetism of the specimen mounting has to be added to each intercept on the $1/(R+a+b)$ axis. This is 15.2 for (a) and (b), 9.0 for (c), 9.9 for (d) and 9.6 for (e). The turning down of the curves (d) and (e) corresponds to the falling off of the susceptibility as the field approaches the critical field.

net moment of the specimen and its surrounding coil is zero, this zero moment is obtained effectively as the sum of slightly positive and slightly negative contributions (due to the detailed geometry of the arrangement), and so the galvanometer gets impulses first one way and then the other.

The exact "profile" of the double kick corresponding to the balance point is difficult to assess, and it was found that the best accuracy could be obtained by interpolation, noting the galvanometer deflections for several values of R definitely below and above the balance value, and plotting the deflections against $1/(R+a+b)$. To reduce the effect of zero instability several readings were taken for each value of R and an average taken. A typical graph is shown in figure 3, and it will be seen that the value of $1/(R+a+b)$ corresponding to zero deflection can be assessed with an accuracy of a few parts in a thousand. The precision increases of course as the magnetic moment, and hence the uncompensated galvanometer deflections, become larger, though a limit is reached when the zero instability begins to grow with the applied field.

As can be seen from equation (6) our procedure finally gives a value of

$\chi V k / A a$, but since k and A could be determined with an accuracy only of order 1% and V only with an accuracy of order 5% (see below), the method is not suitable to give absolute values of χ with sufficient accuracy to measure the small deviation from χ_0 (see equation (3)). Moreover, equation (6) requires slight correction, if, as was the case in most of our measurements, the specimen is appreciably longer than the 5-turn coil; the sensitivity of the double coil method falls off if the specimen does not move exactly from the centre of one coil to the centre of the other, and since evidently this cannot be true for all points of a long specimen, the specimen and compensating coil produce different effects per unit magnetic moment if they occupy different lengths. This correction can be estimated by auxiliary experiments and amounts to nearly 10% for the longest specimens. Fortunately all these uncertainties are irrelevant if we are interested only in relative changes of χ with temperature, for $\Delta\chi/\chi$ does not involve the calibration constants at all, provided that the temperature changes do not alter any of the constants. The only possibility of a change of calibration constants with temperature arose from possible slight movement of the Dewar vessel in which the coils were mounted as the helium vapour pressure was changed; such movement would have altered the relative positions of specimen and coils and hence the calibration constants, but auxiliary measurements showed that this effect was negligible. Thus the values of $\Delta\chi/\chi$ could be measured with an accuracy of a few parts in a thousand.

Since the measured value of χ includes the magnetic effects of the specimen mounting (mostly glass and distrene), it was essential to determine these separately, in case they too varied with temperature. This was done by measuring in a

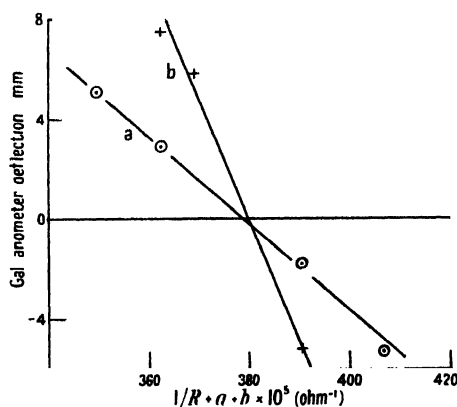


Figure 3. Illustrating interpolation of $1/(R+a+b)$ to zero galvanometer deflection. H5, 4.044 k. (a) solenoid current 20 ma., (b) 50 ma. The interpolated values appear as two of the points on (d) of figure 2

field high enough to destroy superconductivity completely, and the residual effect was found to be paramagnetic, with an effective susceptibility (i.e. per unit volume of the superconductor) of the order of a few per cent of that due to the superconductor. It did indeed vary with temperature, roughly according to Curie's law, so that this correction was quite important; it has been made in all the results quoted below.

(ii) *Temperature Measurement*

All temperatures quoted were deduced from the vapour pressure of the helium bath by means of the 1932 Leiden scale.

(iii) *The Specimens*

The specimens were prepared by filling fine capillaries with spectroscopically pure mercury or tin. In order to obtain enough total volume a large number of these filled capillaries were bundled together so that they were parallel to each other; and separated from each other by about 10 radii. In one case (H2) the necessary volume was obtained by winding a long length of capillary into a spiral with the turns well separated. In the case of the spiral (H2) and of one of the bundles (S5) the field was perpendicular to the length of the superconductor. It should be noted that with this arrangement χ_0 is $1/2\pi$ instead of $1/4\pi$, but equation (3) is unaffected. The gain in sensitivity is offset by the fact that fields only half as large can be used before superconductivity begins to be destroyed. Full magnetization curves of H2 and S5 were taken for the purpose of studying the intermediate state; the results of these measurements are discussed elsewhere (Désirant and Shoenberg 1948). In the case of the other two specimens (H5 and H6) the field was parallel to the cylinder lengths. Details of the specimens are summarized in table 1. The radii quoted were worked out from the values of $\chi V/k_A a$ deduced from the susceptibility measurements after making due allowance for the various corrections mentioned above, and assuming an approximate value of λ in using equation (3) for χ . This was considered more accurate than direct measurement, since measurements with a travelling microscope were difficult, and weighing was also difficult in view of the large numbers of capillaries and the small masses involved. The radii deduced in this way did in fact agree with those estimated by travelling microscope well within the limits of error of the latter method. The microscope readings were useful in indicating the order of magnitude of the variation of radii between the individual capillaries; the estimated mean deviation was 7% of the radius for H5 and H6, and about 20% for S5. Strictly speaking the radius deduced from V is $(\bar{r}^3)^{1/3}$ while that required in deducing $\Delta\lambda$ from the experiment is \bar{r}^2/\bar{r} , which is slightly different if the mean deviation of r is taken into account. Assuming a Gaussian distribution of the radii, it is easily shown that the difference is negligible for H5 and H6, but amounts to 3% for S5. The corrected value of r for S5 is shown in table 2.

§ 3. EXPERIMENTAL RESULTS

For each specimen measurements were made at several temperatures from close to the transition temperature (T_c) down to about 2.1°K. , and from the changes of $1/(R+a+b)$ which is proportional to χ , values of $\lambda(T) - \lambda(2.1^\circ \text{K.}) = \Delta\lambda$ were deduced. The exact rather than the approximate form of equation (4) was used for this purpose, using our final estimate of $\lambda(2.1^\circ \text{K.})$ to correct χ to χ_0

Table 1

Specimen	Radius r cm. $\times 10^4$	Description	T K.	$1/(R+a+b)$ $\times 10^5$	$\Delta\lambda$ cm. $\times 10^6$	$\Delta\lambda/\lambda(2.1)$ from colloids
H2	22.8	Pyrex capillary spiral filled with with mercury. About 6.4 mg. HS.10 763	2.14	303.7 \pm 0.5		
			3.691	302.1 \pm 0.5	6.1 \pm 3	0.52
			3.968	300.6 \pm 0.5	11.9 \pm 3	1.25
			4.046	300.6 \pm 1.0	11.9 \pm 6	1.83
			4.085	296.9 \pm 1.5	25.5 \pm 8	2.34
H5	15.55	Bundle of 98 pyrex capillaries. Mean length 18.3 mm. About 19 mg HS 11 023.	2.16	396.4 \pm 0.5		
			3.928	392.7 \pm 0.5	7.2 \pm 2	1.03
			4.000	391.0 \pm 0.5	10.5 \pm 2	1.45
			4.088	387.4 \pm 1.0	17.5 \pm 3	2.38
			1.72	396.9 \pm 0.5		
			(2.14)	(396.8 \pm 0.5)		
			3.699	393.8 \pm 1.0	5.9 \pm 3	0.52
			4.046	390.8 \pm 0.5	11.6 \pm 2	1.82
			4.046	391.8 \pm 1.0	9.7 \pm 3	1.82
			4.127	384.6 \pm 1.0 -15.0 }	23.8 \pm 31 -3 }	3.97
			(2.14)	(394.8 \pm 0.5)		
			3.19	394.1 \pm 0.5		
			4.044	388.5 \pm 0.5	12.3 \pm 2	1.82
			4.127	379.5 \pm 0.5	29.8 \pm 2	3.97
H6	9.03	Bundle of 198 pyrex capillaries. Mean length 18.6 mm. About 13 mg. HS.11 023	4.045	266.7 \pm 0.5	See below	
			4.129	255.8 \pm 5.0 - 0.5 }		
			(2.14)	277.5 \pm 0.5		
			3.19	276.6 \pm 0.5		
			4.000	270.1 \pm 1.0	11.8 \pm 3	1.45
			4.048	269.4 \pm 1.0	12.9 \pm 3	1.87
			4.085	265.0 \pm 0.5	20.0 \pm 2	2.33
			(4.129)	258.7 \pm 6.0 - 1.0 }	30.1 \pm 2 - 11 }	4.04

Notes to table 1.

(i) The different series of values represent experiments on different occasions; the slightly different values of $1/(R + a + b)$ at 2.1° K. in different series are due to slight differences in the exact geometry of the specimen relative to the measuring apparatus. The values of $1/(R + a + b)$ for different specimens are not exactly proportional to the masses, since different coils c were used on different occasions.

(ii) Of the two readings at 4.046 K. on H5, the first was taken before the low temperature reading and the second after.

(iii) The first of the readings on H5 at 4.127° K. has such large limits of error that it has been ignored in the calculations.

(iv) The first series of readings on H6 was accidentally cut short, but the difference between the 4.045 K. and 4.129° K. readings has been transferred to the next series to give an estimate of $\Delta\lambda$ at 4.129° K.

Table 2

Specimen	Radius r cm. $\times 10^4$	Description	T K.	$1/(R + a + b)$ $\times 10^5$	$\Delta\lambda$ cm. $\times 10^6$	$\Delta\lambda$ cm. $\times 10^6$ (Pippard)	$\Delta\lambda$ cm. $\times 10^6$ (Laurmann and Shoenberg)
S5	17.8	209 pyrex capillaries in 5 layers, field applied transversely. Mean length 7.9 mm. About 11 mg. JM.12 966	2.16	974.0 \pm 1			
			3.009	971.4 \pm 2	2.3 \pm 2	1.2	2.4
			3.555	965.1 \pm 1	8.1 \pm 1	6.2	7.5
			3.598	963.2 \pm 2	9.9 \pm 2	7.3	9.2
			3.649	958.6 \pm 2	14.0 \pm 2	9.9	12.3
			3.667	949 \pm 8	22.7 \pm 7	11.2	13.9

by means of equation (3) *. The results of the various series of measurements for mercury and tin are collected in tables 1 and 2; the errors indicated[†] as \pm are rough estimates of probable error based on the individual extrapolations (similar to those of figure 2) which yield the appropriate values of $1/(R+a+b)$. In a few cases it was not possible to take accurate readings at 2.1°K. , and in such cases an estimated value of $1/(R+a+b)$ at 2.1°K. has been put in the table for the purpose of calculating $\Delta\lambda$. These estimates (shown in brackets in table 1) were made from the value at the nearest temperature used, on the basis of the colloid results assuming an approximate absolute value of λ , or by interpolation, and the error involved is small, since the λ - T curve is nearly flat over a wide range (e.g. below 3.7°K. for mercury).

The usual order of experiment was to measure at successively lower temperatures, and it was necessary to make sure that the observed changes of $1/(R+a+b)$ were genuinely due to changes of penetration depth rather than to some stray effect which was associated with the passage of time (e.g. falling of the liquid helium level in the Dewar vessel). For this purpose, in one experiment (on H5) the measurements at 4.048°K. were repeated after those at 2.1°K. had been taken; as can be seen from table 1, no significant difference was found, thus confirming that the observed changes were indeed only due to the temperature changes.

In order to derive an absolute value of λ for mercury our results must now be compared with those from the colloid measurements (Shoenberg 1940). From the latter we have derived values of $\Delta\lambda/\lambda(2.1^\circ\text{K.})$ at each temperature used in the present experiments, and these are shown in the last column of table 1. The comparison assumes[†] that T_c for our specimens was the same (4.17°K.) as for the colloid specimen, but actually this assumption was not very carefully verified, since its importance was realized only later. From such few measurements of critical field as were made, however, it is unlikely that T_c differs from 4.17°K. by more than 0.01°K. The simplest method of deriving $\lambda(2.1^\circ\text{K.})$ is to plot the observed values of $\Delta\lambda$ from the present experiments against those of $\Delta\lambda/\lambda(2.1^\circ\text{K.})$ from the colloid results. This has been done in figure 4, and it will be seen that within the limits of experimental error the relation is a linear one. This confirms the validity of equation (4), since there is no systematic difference between the points for specimens of different radii, and also the consistency of the present results with the colloid measurements.

The slope of the best straight line through the points of figure 4 gives $\lambda(2.1^\circ\text{K.})$ in absolute measure, and a least squares calculation using weights inversely proportional to the errors shown in table 1 gave $\lambda(2.1^\circ\text{K.}) = 7.9 \times 10^{-6}\text{ cm.}$ with a standard deviation of $0.3 \times 10^{-6}\text{ cm.}$ Extrapolation to absolute zero by means of the colloid results then gives $\lambda(0^\circ\text{K.}) = 7.6 \times 10^{-6}\text{ cm.}$ This estimate may well be in error by a greater amount than indicated by the small standard deviation, since the calculation cannot, of course, allow for possible systematic errors in the colloid results, for possible differences in T_c between the present specimens and the colloid (which might introduce a systematic error) and for unsuspected systematic errors in the present experiments. In figure 5 the values of $\Delta\lambda$ are

* In the case of tin, $\lambda(2.1^\circ\text{K.})$ was assumed to be $5 \times 10^{-6}\text{ cm.}$ for this purpose but since the correction is only about $\frac{1}{4}\%$, the uncertainty in this assumption is unimportant.

† It is probable that for small changes of T_c , λ remains the same function of $T_c - T$, so if T_c is different for two specimens, the observed values of T must be adjusted in order to compare values of λ .

shown plotted against $\Delta T = T_c - T$, (assuming $T_c = 4.17^\circ \text{K.}$) together with a graph of the colloid results assuming $\lambda(2.1^\circ \text{K.}) = 7.9 \times 10^{-6} \text{ cm.}$ The broken-line curve comes from the results of Appleyard *et al.* (1939) which are discussed below.

The results for tin cannot be treated in the same way as those for mercury, since no colloid measurements are as yet available. They may, however, be compared with those of Pippard (1947) and with those of Laurmann and Shoenberg (1947). In making such a comparison due allowance must be made for the fact that S5 had a transition temperature $T_c = 3.738 \text{ K.}$ while the specimens of the other two experiments had $T_c = 3.711 \text{ K.}$, for each temperature of measurement of S5 we therefore show in table 2 the value of $\Delta\lambda$ obtained in each of the other two experiments at a temperature 0.028 K. lower. It will be seen that except at the highest temperature (where owing to the limited range of fields available the accuracy was rather poor), the results agree fairly well with those of Laurmann and Shoenberg, but are a little higher than those of Pippard.

It is possible that this discrepancy is associated with anisotropic effects, but it will be convenient to defer a detailed discussion of this point to paper II.

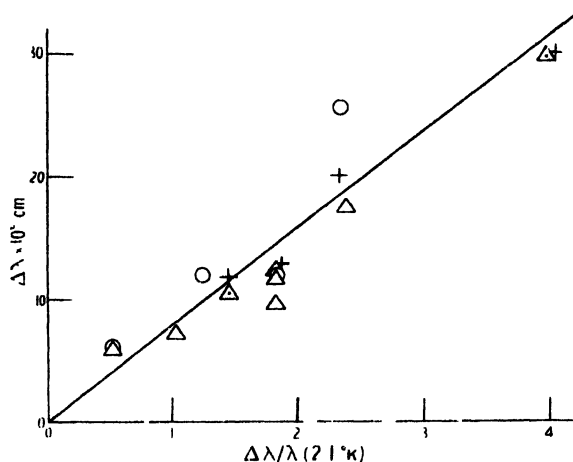


Figure 4. Relation between $\Delta\lambda$ from present work and $\Delta\lambda/\lambda$ deduced from colloid experiments. (○) H2, (△) H5, (+) H6.

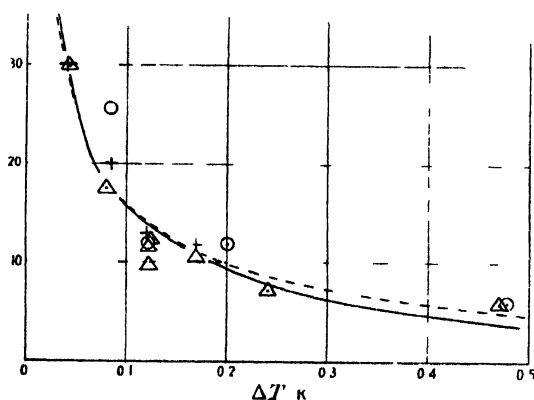


Figure 5. Relation between $\Delta\lambda$ and ΔT (assuming $T_c = 4.17 \text{ K.}$). (○) H2, (△) H5, (+) H6. Full curve deduced from colloids assuming $\lambda(2.1 \text{ K.}) = 7.9 \times 10^{-6} \text{ cm.}$ Broken curve from thin film measurements of critical field explained in text.

§ 4. RELATION BETWEEN PENETRATION DEPTH AND CRITICAL FIELD

As has already been mentioned, information about penetration depth may be obtained also from the measurements of critical fields of thin mercury films by Appleyard *et al.* (1939). The interpretation is complicated by the necessity of allowing for a possible difference between the surface tensions α_n and α_s at a boundary between an insulator and the normal and superconducting phases respectively. If the film thickness is not too small, Ginsburg (1945) has shown thermodynamically that the critical field h of a film of thickness $2r$ is given by

$$h/H_0 = 1 + (\lambda + \beta)/r, \quad \dots\dots (7)$$

where
$$\beta = 8\pi(\alpha_n - \alpha_s)/H_c^2 \dots\dots\dots (8)$$

and H_c is the critical field of the bulk metal. Equation (7) is valid, however, only if $(\lambda + \beta) \ll r$, while most of the thin films of Appleyard *et al.* had $r \lesssim \lambda + \beta$. For such thin films the thermodynamic theory can be developed only by assuming a detailed form of penetration law, and Ginsburg in fact assumed the exact validity of the Londons' law (1); he assumed further that the concept of a surface tension still applies even when $r \lesssim \lambda + \beta$, which would seem to require some justification. On the basis of these assumptions Ginsburg calculated from the thin film data a curve of λ against T which disagrees seriously with that of figure 5. The lack of agreement suggests that Ginsburg's assumptions are probably unjustified.

The analysis of their results made by Appleyard *et al.* suggested that within experimental accuracy h/H_c was a function of a single temperature dependent parameter divided by r . They identified this parameter with the penetration depth λ , but in view of the complications just discussed, it is safer to suppose that it has some other meaning and we shall denote it by λ' ; for thick films, for instance, λ' should be the same as $\lambda + \beta$. If we assume $\lambda'(2.1^\circ \text{K.}) = 10.8 \times 10^{-6} \text{ cm.}^*$, we find that the curve of $\Delta\lambda' = \lambda'(T) - \lambda'(2.1^\circ \text{K.})$ against ΔT (the broken line of figure 5) agrees fairly well with the colloid curve of $\Delta\lambda$ against ΔT . This suggests that $\lambda' - \lambda$ is fairly independent of temperature, and if, speculatively, we identify $\lambda' - \lambda$ with β , this result might be taken to mean that β is about $3 \times 10^{-6} \text{ cm.}$ at high and low temperatures, rising to about $4 \times 10^{-6} \text{ cm.}$ at 3.7°K.

§ 5. POSSIBLE FIELD DEPENDENCE OF λ

An interesting possibility, first suggested by Ginsburg (1947), is that the penetration depth may depend on the strength of the applied magnetic field. Most of our measurements were made at low fields, and up to about $\frac{1}{2}H_c$ there was certainly no appreciable effect of this kind. A few measurements were, however, made at higher fields and we did find that χ began to fall off appreciably; some typical curves of $\frac{1}{2}r\Delta\chi/\chi$ against H/H_c for H5 and H6 at 4.05°K. and 2.16°K. are shown in figure 6, and similar effects were found in longitudinal tin specimens. If these changes of χ are interpreted as due to changes of λ with field, the values of $\frac{1}{2}r\Delta\chi/\chi$ would represent those of $\lambda(H) - \lambda(0)$; since this quantity is greater for the thinner specimen H6 than for H5, this does not seem to be a satisfactory interpretation, though it should be emphasized that these measurements were not very accurate, and it is just possible that the differences between H5 and H6 are due to experimental error. Probably the observed effects are at least partly due to some

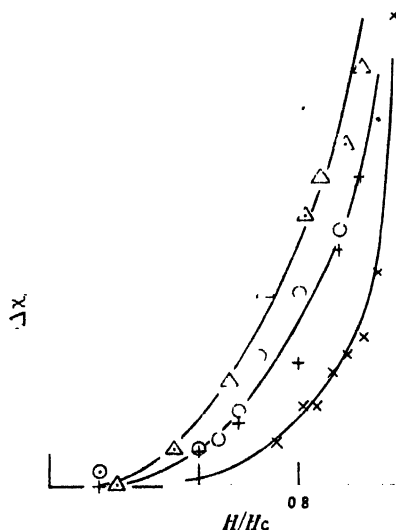


Figure 6. Relation between $\frac{1}{2}r\Delta\chi/\chi$ and H/H_c .
 — H5 4.046°K. — H5 2.14°K.
 ○ H5 4.047°K. , on another occasion.
 △ H6 4.045°K.

* Appleyard *et al.* gave $\lambda(2.1^\circ \text{K.}) = 12 \times 10^{-6} \pm 2 \times 10^{-6}$ based on the values of h/H_c for their thickest films, so our assumption is not inconsistent with their results.

slight destruction of superconductivity starting at irregularities or at the ends of the cylinders. A similar effect was found by Laurmann and Shoenberg in their repetition of the Casimir experiment, and there the apparent increase of λ was always associated with the appearance of resistive losses. Although evidently this question requires further investigation, we may say from the present experiments that there is no appreciable increase of λ up to $\frac{1}{2}H_c$, and that for higher fields an increase cannot be excluded, though there is no proof of its existence.

It may be noted that if in fact λ does increase with field, this would invalidate the argument on which equation (7) is based, and would further complicate the interpretation of critical field measurements, such as those of Appleyard *et al.*

ACKNOWLEDGMENTS

We wish to thank Mr. A. B. Pippard for many helpful suggestions and discussions, and Mr. J. M. Lock for help in preparing the diagrams.

REFERENCES

- APPLEYARD, E. T. S., BRISTOW, J. R., LONDON, H., and MISENER, A. D., 1939, *Proc. Roy. Soc. A*, **172**, 540.
 CASIMIR, H. B. G., 1940, *Physica*, **7**, 887.
 DÉSIRANT, M., and SHOENBERG, D., 1947, *Nature Lond.*, **159**, 201; 1948, *Proc. Roy. Soc. A*, in press.
 GINSBURG, V., 1945, *J. Phys. U.S.S.R.*, **9**, 305; 1947, *Ibid.* **11**, 93.
 LAURMANN, E., and SHOENBERG, D., 1947, *Nature, Lond.*, **160**, 747.
 LONDON, F. and H., 1935, *Proc. Roy. Soc. A*, **149**, 71.
 PIPPARD, A. B., 1947, *Nature, Lond.*, **159**, 434.
 SHOENBERG, D., 1940, *Proc. Roy. Soc. A*, **175**, 49; 1947, *International Conference on Fundamental Particles and Low Temperatures*, Vol. II (Physical Society), p. 85.

The Rate of Growth of Current and the Behaviour of the Cathode Spot in Transient Arc Discharges

BY K. D. FROOME

1851 Senior Student, Department of Physics, Imperial College, London

Communicated by G. P. Thomson; MS. received 25 November 1947

ABSTRACT. The factors are discussed which control the initial rate of increase of current in the arc-like discharge obtained when a condenser of several microfarads is discharged through a small tube containing rarefied gas. It is suggested that at higher pressures ionization probably proceeds "thermally", but that at lower pressures ionization takes place by direct electron collision, the cathode spot having a negligible retarding effect except possibly at very low pressures. The current density of emission from the cathode spot is at all times enormous, appearing sometimes greater than 10^6 amperes per cm^2 and very seldom less than 10^5 amperes per cm^2 even for a discharge prolonged until it has the appearance of a normal arc. It is suggested, therefore, that the value (Druyvesteyn and Penning 1940) of the emission current density of about 10^4 amperes per cm^2 usually accepted for normal vacuum arcs is in error (Froome 1946, 1947). The rapid movement of the cathode spot and rate of growth of current exclude the possibility of the cathode emission being derived thermionically, while the very high current densities observed favour field emission.

§ 1. INTRODUCTION

ALTHOUGH the transient arc obtained by condenser discharge through a tube filled with gas at reduced pressure is widely used for high speed photography (for list of references see Henry 1944), a systematic study of the factors determining the growth of current seems to have been neglected. The discharge will be termed for convenience an "arc" since it needs, in common with a normal arc, a cathode spot for its maintenance, although it differs considerably in other respects from a normal arc, especially during its initial stages. In this work a condenser of a few microfarads, charged to some 360 volts, is connected directly across the tube, the discharge being started by applying a brief high-tension pulse to an electrode wrapped round the outside of the tube, adjacent the cathode. The discharge tube usually has an electrode spacing of about 18 mm. The paper consists of two parts: firstly the study of current rise when the impedance of the leads between condenser and tube is a minimum except for a small inductance in series with a small resistance included in the cathode lead for measuring purposes; secondly the study of the cathode spot by a simple Kerr cell technique, where sometimes a fairly large resistance has been added to the leads in order that the transition to a normal arc can be studied.

§ 2. THE RATE OF CURRENT RISE

Discussion will be limited to studies of times of the order of 10^{-7} second, since the effects of possible cathode vaporization and of impurities arising from this seem negligible in such times. It is not proposed, in this section, to trace the current rise above 100 amperes or so, although the current may rise to over 1000 amperes in a microsecond if the external impedances are kept very low.

To measure such times a cathode-ray tube is used on a repetitive system so that a fair sensitivity of Y-shift is obtained. This should not be less than 1 cm. shift for 30 v. applied; otherwise too much of the condenser voltage will be taken up in the measuring circuit. The trace is photographed with an F3.5 lens. With the tubes and condenser voltages used there is a variable delay of the order of $\frac{1}{10}$ to 1 microsecond, occasionally as much as 100 microseconds, between the application of the triggering pulse and the start of the arc. Further, if a cathode spot does not form during this time, a glow discharge may strike, but not an arc. The probability of the formation of an unwanted glow discharge can be reduced by making the electrode area small, and the cathode of metal of low work function, or of mercury; this does not upset the discharge if an arc forms, for in this case ionization follows a narrow intense channel. Because of this delay, an independent linear time base is useless for measuring durations of 10^{-7} sec. or less unless the photographic writing speed of the cathode-ray tube is such that the current at any time in the discharge tube can be measured on a single trace. This demands a cathode-ray tube operating at an anode voltage of at least 5000, and a faster lens, implying too great a sacrifice of deflection sensitivity. Thus the method adopted was to allow the voltage (or a fraction of it) across the discharge tube to supply the X-deflection, and to record the rate of growth of current, and the current, on the Y-shift. If a double-beam tube is used these can be recorded simultaneously, but if a single beam tube is used, then two exposures must be superimposed.

This method was found to be very satisfactory, for once an arc starts, the relation between the potential difference across the discharge tube and the current or rate of

growth of current in it is always the same, so that the trace obtained, even after several thousand superimposed flashes, is quite sharp. The cathode-ray tube beam is turned to full brilliancy for a time a little longer than the duration of the discharge plus the variable delay already mentioned.

Figure 1 shows a typical discharge tube and the measuring circuit employed for these current rise experiments. L is a small inductance (about 10^{-7} henry) in series with a small resistance R (about 1 ohm) in the cathode lead of the discharge tube. The junction of these, E , is earthed, so that the potential across L or R can be applied to the Y plate of the cathode-ray tube, and thus di/dt and i can be found if the values of L and R are known. Hence the area under a graph of dt/di against i gives a time scale, since

$$t = \int (dt/di) di.$$

The minimum detectable current is about 1 amp. and for convenience we shall measure time from this point. This method will resolve 10^{-8} second.

Figure 2 shows the recurrent single-sweep linear time-base and pulse generator, the time base portion being used in the second half of this investigation. It provides the triggering pulse to the discharge tube, as well as switching on the cathode-ray tube beam to full intensity for a time equal to that of the linear sweep. This can be varied between 3 and $10\,000\ \mu\text{sec.}$ in 8 steps. It is used on a VCR.97 with anode voltage of 2500, the most sensitive plates being used for Y deflection. The linear sweep is used for the set up of figure 1.

In figure 2, the thyatron V1, is used to trigger the whole sequence of events. It can be used as a slow relaxation oscillator, or triggered singly by a pulse applied through C2, or synchronized to a sub-multiple of the mains frequency by closing S1. This latter alternative is used when a large number of recurrent traces is needed, R8, R2 being adjusted for a repetition speed of about 10 per sec. Single triggering is used for the Kerr cell shutter investigations described later. When V1 has "fired", it triggers the thyatron V3, which discharges one of the time-base condensers C15-C22 at constant current through the beam tetrode V2, thus providing a single high-speed linear sweep for supplying the cathode-ray tube X-deflection if required. V2 should have a screen potential of about 400 v. in order that this may be large compared with the potential drop across V3 when in the conducting state. After V3 had fired, it was found that it would not recover until one of the tetrode screen decoupling condensers C6-C8 had been discharged. It was also found that for a GTIC thyatron the resistance R21 and small condensers C13 were needed to assist recovery. These should be as small as possible. It is thus apparent that the tetrode screen decoupling condenser C6-C8 must be of such a value as not to discharge appreciably during the linear part of the sweep—but must do so before V1 produces a further triggering pulse to V3.

During the time of linear discharge a small negative square-wave pulse appears across R22; this is amplified by V4, and the resulting positive square-wave pulse

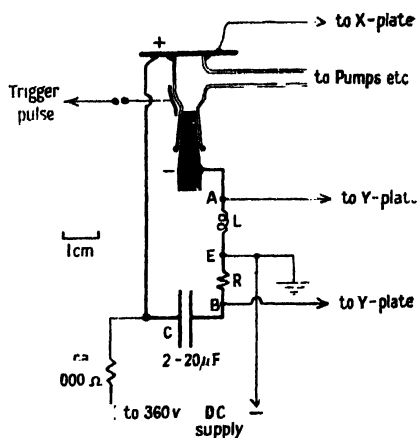


Figure 1.

fed to the cathode-ray tube grid through C31, thus turning this tube to full brilliance for the duration of the linear discharge. It can also be used to trigger the Kerr cell circuit described later. The thyatron V5 can be triggered after V3 through the variable delay network D2 by this positive pulse from V4, or it can be triggered before V3, the triggering of V3 then being delayed by the variable lag D1, depending on the position of the switch S2. When V5 fires, the condenser C33 is discharged through the primary of a small induction coil, thus producing the high

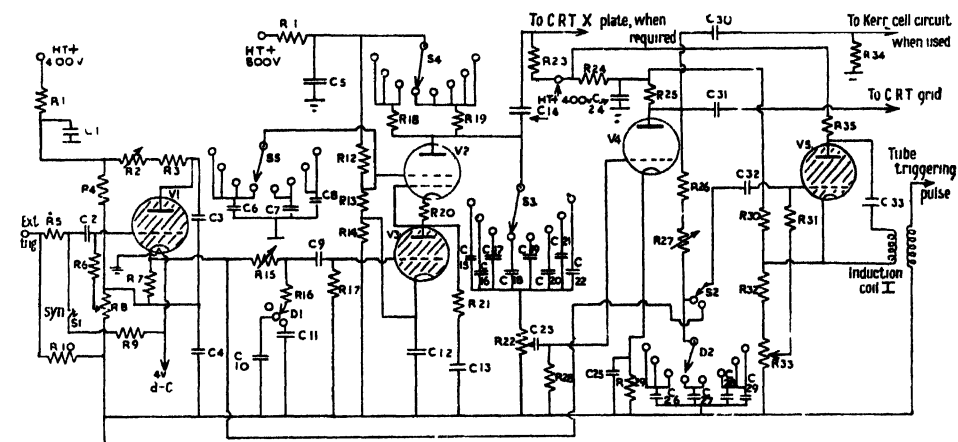


Figure 2.

Resistances (R)		Condensers (C)		Valves
R1, 11, 14, 15, 17, 24, 25, 33	10 000 Ω	C1, 24	8 $\mu\text{F.}$	V1, V3, V5 Type 6T1C
2	250,000 Ω	2, 3, 6, 9, 14, 19, 23, 29, 30, 31, 32	0.1 $\mu\text{F.}$	2 Type 6L6
3, 6	200,000 Ω	4, 7	0.5 $\mu\text{F.}$	4 Type 6J5
4	89,000 Ω	5, 8, 22	4 $\mu\text{F.}$	
5, 12, 13, 19, 31	100,000 Ω	10, 15, 27	0.001 $\mu\text{F.}$	
7, 16	1,000 Ω	11, 17, 28	0.01 $\mu\text{F.}$	
8	5,000 Ω	12, 25	20 $\mu\text{F.}$	
9	22,000 Ω	13	0.005 $\mu\text{F.}$	
10, 23, 28, 34	1 M Ω	16	0.003 $\mu\text{F.}$	
18	300,000 Ω	18	0.03 $\mu\text{F.}$	
20, 21	30 Ω	20, 33	0.25 $\mu\text{F.}$	
22	100 Ω	21	1 $\mu\text{F.}$	
26	50,000 Ω	26	0.0002 $\mu\text{F.}$	
27	3 M Ω			
29	500 Ω			
30	80,000 Ω			
32	4,200 Ω			
35	69,000 Ω			

tension pulse which finally triggers the discharge tube under investigation. The delay networks D1, D2, are needed to ensure that the discharge tube arc starts whilst the cathode-ray tube beam is at full brilliance. The 8-position switches S3, S4, S5, D2 are ganged.

Figures 3, 4, 5 show some experimental results for a tube of 18.5-mm. electrode spacing, connected across a condenser of 2 $\mu\text{F.}$, initially charged to 360 v., at various pressures and natures of gas filling.

It is seen that at about 2 mm. Hg pressure there is a marked difference in the rates of current rise for air, helium and argon filling. With helium the current rises to 50 amp. in about 10^{-7} sec., whereas with air it takes about $2\frac{1}{2}$ times as long.

Even so, the speed with which the current increases is very great for both, suggesting that whatever is the emission process operating at the cathode, it can build up extremely rapidly. This emitting area or "cathode spot" will be discussed next, before attempting theoretical estimates of the current rise to be expected. It has not been found that the nature of the cathode has a significant effect on the current rise, provided measurements (such as the above) are made before extensive vaporization takes place from it. These results of figures 3-5 have been obtained

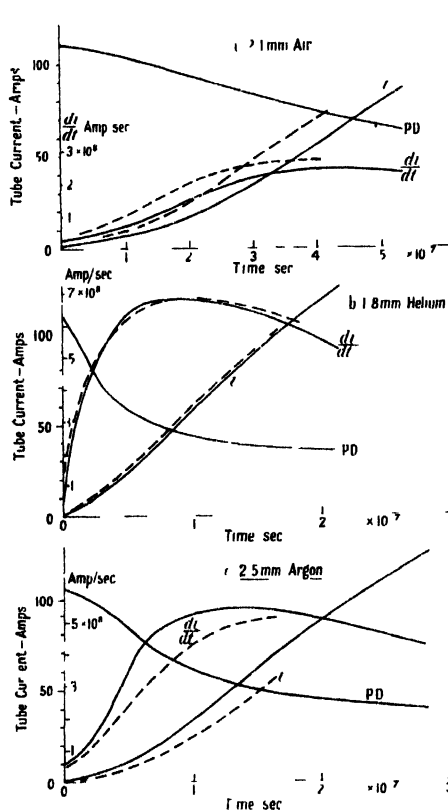
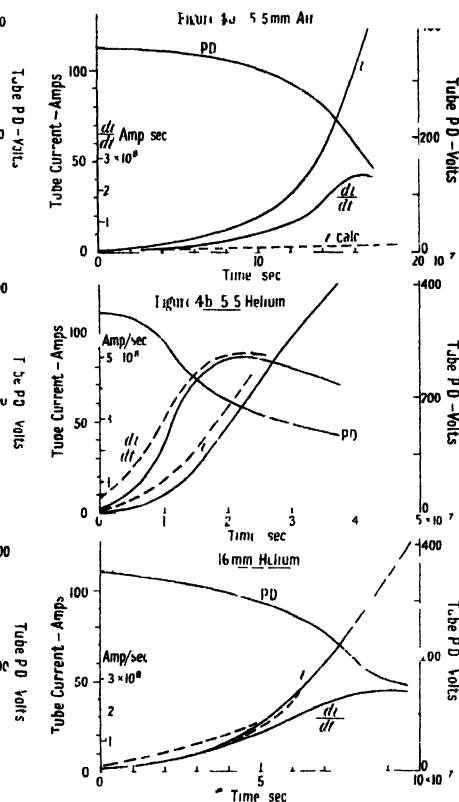


Figure 3.



Figures 4 & 5.

Figure 4: 5.5 mm. Figure 5: 16 mm

with a sodium cathode with enough lead added to render it sufficiently hard for machining on a lathe. Before use the protective layer of oxide which forms on the surface is removed, the cathode quickly inserted into the discharge tube, and the discharge run for some thousand flashes so that the cathode spot can further clean the cathode under it.

§ 3. THE BEHAVIOUR AND SIZE OF THE CATHODE SPOT

A study of the cathode spot during a transient condenser discharge is important for two reasons. Firstly, the nature of the cathode spot is liable to influence the rate of growth of current. Secondly, if the arc is prolonged by addition of a resistance of several ohms in the lead between condenser and discharge tube, until it has the characteristics of a normal vacuum arc, then the small size of the tube

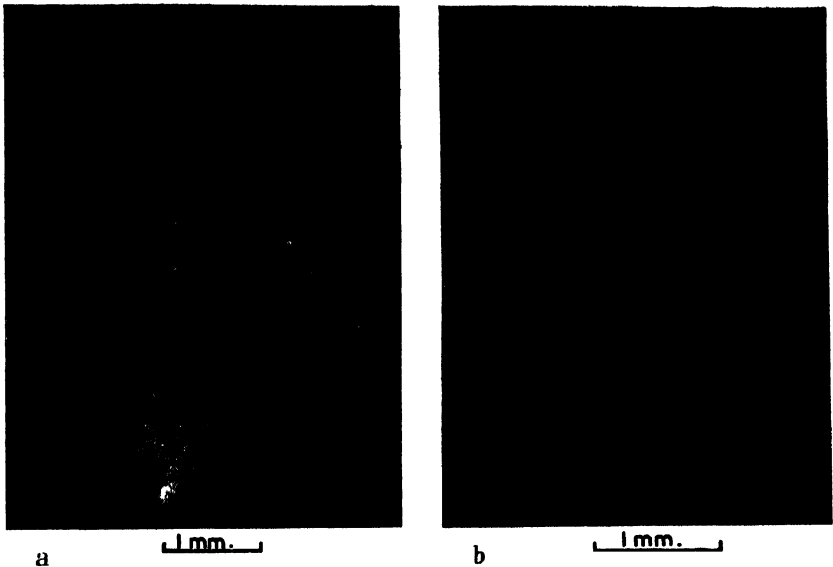


Figure 8

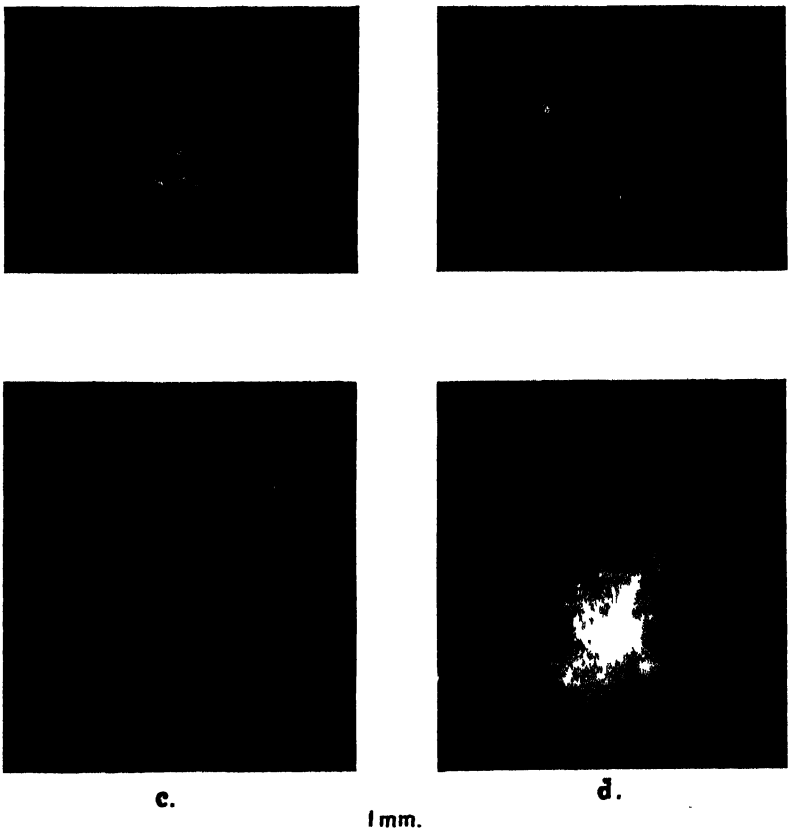


Figure 9.



Figure 11a.

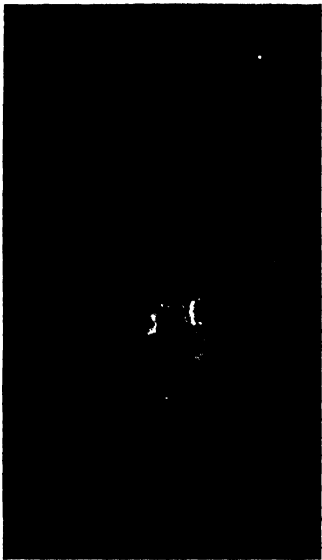


Figure 11b.



Figure 12a.

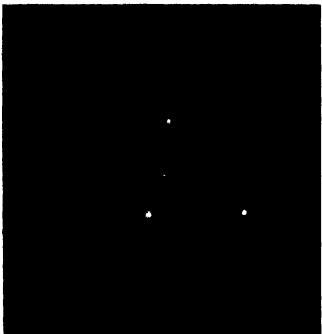


Figure 12b.

1 mm. —

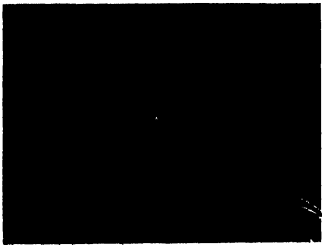


Figure 13.

enables the cathode spot to be studied under much better conditions than in a tube designed for continuous running at comparable currents.

A Kerr cell shutter has been used to take low power photomicrographs of the discharge-tube cathode at any stage during the condenser discharge. These are taken singly, on ciné film, and subsequently enlarged until the final overall magnification is about 20 diameters. The ciné film technique is of value when short exposures of moving objects are needed (Froome and Jarrett 1943, 1944).

Figure 6 shows the circuit diagram of the Kerr cell shutter, which is triggered via a delay circuit by the positive pulse from the amplifier valve V4. The optical arrangement for photographing a tube with a mercury cathode is shown in figure 7. In this particular tube the distance from the top of the cathode to the optical flat covering the tube was about 25 mm.

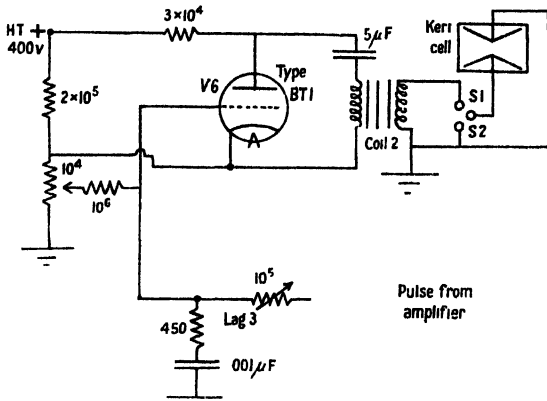


Figure 6.

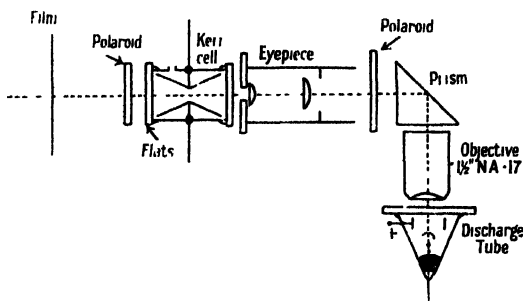


Figure 7.

When the thyatron V6 (figure 6) is triggered, the 5-μF. condenser is discharged into the primary of induction coil 2. The resulting high tension developed across the secondary then sparks over the spark gap S1, opening the Kerr cell shutter until S2 sparks over, when the shutter closes once more. The spark gaps S1, S2, must be arranged so that light from S1 adequately illuminates S2. Both S1 and S2 can be varied by screw adjustments. By judiciously varying S1 and S2, exposures can be obtained ranging from about $\frac{1}{10}$ to 40 μsec. When it is desired to use this apparatus, the cathode-ray tube X-sweep is provided by the linear time-base already described, the length of exposure, time of exposure after initiation of discharge tube arc, and arc current being observed on the screen, either visually or by photography. If exposures less than about 3 μsec. are required, only visual observation can be used, since there is a variable delay of at least 1 μsec. between the application of the triggering pulse to the discharge tube and the development of the arc, even when operating under optimum conditions. There is a similar degree of uncertainty in the sparking over of the gap S1 of figure 6. If longer exposures are used (as when studying prolonged arcs) the cathode-ray tube can be photographed by repeating the whole process a sufficient number of times. In this case, of course, the film of the Kerr cell shutter is exposed for one condenser discharge only. There is also a variation in the exposures given by the

Kerr cell shutter for any given setting of the gaps S1 and S2, but the exposure actually obtained on the ciné film can always be observed sufficiently accurately on the cathode-ray tube screen. In practice, exposures have usually been made as short as possible consistent with adequate exposure of the ciné film, except when it was desired to show movement of the cathode spot during a fairly long exposure. Some results for various cathodes will now be discussed; these are all characteristic of a large number of similar photographs. It was not found that the pressure of the gas filling of the discharge tube had any significant effect on the appearance of the cathode spot; in all cases in this investigation the tubes have therefore been run at conditions of optimum ease of triggering.

Mercury cathode

Figure 8(a) shows a photomicrograph of the mercury cathode during the whole of one discharge of $20\mu\text{F}$. initially at 300 v. The external impedances between condenser and discharge tube were made a minimum, the L and R of figure 1 being reduced to $3 \times 10^{-8}\Omega$ and 0.05Ω respectively, and the total residual inductance of the condenser and leads about $10^{-7}\Omega$. Under these conditions the discharge lasts some $7\mu\text{sec}$. and the current reaches a peak of 1400 amp. about half-way through this time. It is found that the cathode emitting area always produces a similar, but not identical, photograph. The "spot" usually forms on the mercury surface adjacent to the tube wall, and expands outwards covering the area shown. For different discharges this area is approximately constant in size, but varies in shape. Figure 8(b) shows a Kerr cell picture of an identical discharge with an exposure of about 10^{-7} second, taken with tube current at the maximum. It is now apparent that the area exposed in figure 8(a) is caused by much smaller, rapidly moving emitting areas, the current density of emission from these being, for this stage of the discharge, greater than 10^6 amp/cm^2 . Reducing the maximum current by reducing the initial condenser voltage merely seems to reduce the area of emission. Before discussing this further, we will examine the behaviour of the "spot" when the arc is prolonged to about $90\mu\text{sec}$. (thus reducing the peak current) by adding a resistance of 1.6Ω between condenser and discharge tube. Figure 10 shows the tube current and voltage drop plotted against time for such a discharge. It also shows the positions A, B, C of the Kerr cell exposures corresponding to the photographs in figure 9(a), (b), (c), respectively.

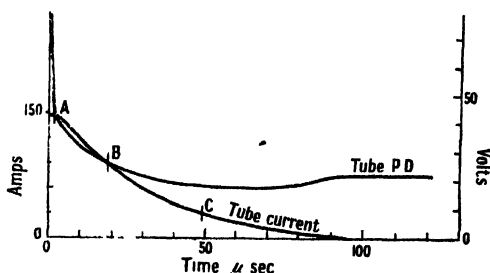


Figure 10.

Figure 9(a) shows the cathode spot for the first $5\mu\text{sec}$. of the discharge. Figure 9(b) shows a longer exposure of $7\mu\text{sec}$. taken about $18\mu\text{sec}$. from the start, and figure 9(c) shows an $8\mu\text{sec}$. exposure taken about $50\mu\text{sec}$. from the start, when the arc has apparently become a "normal arc". Figure 9(d) shows the cathode for the whole of one flash. It is easy to see how this picture is built up from (a)-(c). In figure 9(a) the "spot" appears similar (though on a smaller scale) to that of the minimum impedance discharge of figure 8(a), probably consisting of a very small, nebulous, rapidly moving area. In figure 9(b), the

"spot" has moved further over the mercury surface away from the tube wall, is spreading out, and developing into a number of separate unit-like spots. The fine streaks shown indicate these moving radially outwards from the discharge starting point. In (c), the unit-spot formation is complete, the current being carried by some 15–20 spots each presumably emitting about 2 amp. From this point the motion of the spots is random; they remain in a group, but at relatively large distances apart. The current density of emission is probably at least 10^5 amp/cm². In this case the average diameter of the spots appears to be about $\frac{1}{20}$ mm. After this spot-formation is complete, no further change takes place; the number of spots merely falls with falling current. This suggests that the usually accepted value of the cathode spot emission current density for mercury arcs is at least an order low. Further photographs, with the Kerr cell shutter, of 50 amp. and 100 amp. normal arcs, $\frac{1}{100}$ sec. after their initiation, confirm this (Froome 1946). There can be little doubt that an arc is "normal" after this interval.

Sodium cathode

Figure 11(a) shows the appearance of a freshly scraped sodium cathode for the whole of one minimum impedance discharge, under electrical conditions identical with those obtaining for the discharge shown in figure 8. This photograph shows an unusual, but not rare, occurrence in that the arc struck simultaneously in three cathode "spots". Figure 11(b) shows an exposure of 10^{-7} sec. taken at the maximum of an identical discharge with the arc current at 1400 amp. Again it is found that the emitting area at any instant is only a fraction of the whole area traversed and that the emission current density is in excess of 10^6 amp/cm² at this stage.

The result of prolonging the discharge by the addition of a series resistance of 7 Ω between condenser and tube is seen in figure 12: figure 12(a) shows the appearance of the cathode for the whole of one flash and figure 12(b) shows a 2 μ sec. Kerr cell exposure of the cathode, taken 120 μ sec. from the time of arc initiation, with the arc current at 12 amp. The unit-spot formation illustrated for a mercury cathode in figure 9, also takes place for a sodium cathode; the number of emitting spots has dwindled to three (figure 12(b)). Hence, a current density well in excess of 2×10^5 amp/cm² is found, although the arc appears "normal" at this stage, having a potential drop of some 20 v. and a negative characteristic.

Copper cathode

An important difference was noted between copper cathodes and mercury and sodium cathodes: for mercury or sodium the cathode "spots" were brilliantly visible during the very high current minimum impedance discharges and the early stages of prolonged discharges, whereas the copper cathode "spot" was not, and in fact was often apparently non-existent. It was found, however, that a clean, polished cathode was marked after one such flash in a manner not unlike the photographs of figures 8(a) or 11(a). This suggested that the cathode emission process was not significantly different, but that at the estimated velocity of the emitting area for mercury or sodium (about 10^8 cm/sec. for a 1400-amp. "spot") the copper cathode was not vaporized and thus the "spot" was much less brilliant. By making the conventional supposition that one-tenth the arc current was carried by positive ions, and assuming a cathode fall (see later) of about 20 v., calculation shows that a mercury cathode could be easily vaporized in the time of about

10^{-7} sec. that the cathode spot takes to move on to fresh territory, while copper would not even be heated to red heat. It has been seen, however, that for a prolonged discharge with mercury or sodium cathodes, the unit-spots formed after the arc has become "normal" move more slowly (of the order of 10^3 cm/sec.) and hence it might be expected that for a similar discharge with copper cathode these unit-spots should appear once the copper in their neighbourhood became incandescent or vaporized. This was in fact found: after some $30 \mu\text{sec.}$ very brilliant unit spots became visible. They moved more slowly than for mercury and sodium, and were considerably brighter in appearance, presumably due to incandescence of the copper. Their size was again very small, indicating emission current densities of at least 2×10^5 amp/cm², their numbers falling with decreasing current. Each spot carried about 3–5 amp.

Figure 13 shows these spots with an exposure of $4 \mu\text{sec.}$ with the arc current at 9 amp., $150 \mu\text{sec.}$ after the start of a discharge prolonged as for the sodium cathode spot photographs of figure 12. An approximate treatment for estimating the cathode temperature rise under the spot is as follows:

Assume that the temperature of the surface, initially θ_0 , is instantaneously raised to θ_1 . Then θ_r at a distance x from this surface is given by an equation of the form

$$\frac{\theta_r - \theta_1}{\theta_0 - \theta_1} = \frac{2}{\sqrt{\pi}} \int_0^y \exp(-y'^2) dy',$$

where $y = x \sqrt{2/(ht)}$ and $h = k/\rho s$, k is the thermal conductivity of the cathode material, ρ its density, s its specific heat, and t the time.

Most of the heat will be concentrated in a distance x given by $\frac{1}{2}x(ht/\rho s)^{-1} \simeq 1$. Thus for copper $x = 2\sqrt{(0.9t/9 \times 0.1)} = \frac{2}{3} \times 10^{-3}$ cm. and for mercury $x = 2\sqrt{(0.02t/14 \times 0.03)} = 1.4 \times 10^{-4}$ cm., if $t = 10^{-7}$ sec., the time taken for the spot to move to fresh territory. We now assume that all the energy supplied to the cathode from the time the spot forms on it, will be used in heating the volume of material of depth x under it.

The input energy to this region is about 2×10^6 w/cm² if the cathode fall is 20 v. and the positive ion current is one-tenth the electron current, the electron current density being about 10^6 amp/cm².

Thus the temperature reached by the copper in this time is found to be 75°C. The temperature that would be reached by the mercury if there were no boiling is 770°C.

Conclusions on cathode spots

(i) At all times the emission current density is enormous, ranging between 10^5 to well in excess of 10^6 amp/cm². (ii) The initial rapid movement of the emitting area precludes all possibility of the emission being derived thermionically. (iii) Intense luminosity of the cathode spot is a secondary effect, only caused if the positive ions falling on the cathode (under the "spot") are able to vaporize it or render it incandescent. It seems probable that the mottled effects of figures 8(a) and 11(a) are due to an unevenness in the velocity of the emitting area, as though it were constantly slowing down and gathering speed, the cathode "spot" appearing brighter when moving more slowly. (iv) Previous measurements of cathode spot current density based upon its markings on the cathode over a long period of time are probably unreliable. It may be that Slepian's (1942) "low current density arc" had a very high cathode current density carried by a large number of small,

rapidly moving areas. (v) The very high cathode current density encountered for both high current arcs and arcs prolonged to normality with mercury, sodium, or copper cathodes, favour Langmuir's original theory (1923) that normal arcs with cathodes of these metals have their cathode emission derived by field emission. Current densities of the order of those herein described, make this theory more plausible than hitherto.

§ 4. THEORETICAL IMPLICATIONS

Discussion in this paper is limited to the simplest calculation of rate of growth of current, time is measured from the point when the current has reached 1 amp. When this has occurred there is a conducting path of positive ions and electrons bridging anode and cathode, but the potential difference applied across the tube is still substantially that across the condenser. Inside the discharge tube this potential difference will be formed of three parts, that due to cathode fall, that due to anode fall, and that due to the necessity of maintaining an electric field inside the plasma or ionized gas between anode and cathode. It is well known that the cathode fall for normal arcs is roughly equal to or less than the ionization potential of the gas filling and this assumption will be made for the arcs examined here, for increasing or decreasing the current does not alter the cathode fall of normal arcs appreciably. If, as seems most likely, field emission is the process by which electrons leave the cathode at the cathode spot, this process does not demand a particularly high cathode fall, and the current rise (as far as this process is concerned) will be equal to the speed at which the necessary positive ion space charge sheath can be built up at the cathode. The anode fall arises from the need for a rate of ionization slightly greater just near the anode than in the plasma. The same density of ionization is needed at the anode as in the plasma, but there is a slow drift of positive ions away from the anode, so that the anode fall is needed to make good such losses. Again, if we assume an anode fall equal to the gas ionization potential, adequate ionization can be produced. On these hypotheses we see that some 20–50 v. of the initial overall voltage of approximately 360 are absorbed by the combined anode and cathode falls.

We shall assume that the remainder is applied to the plasma and calculate the rate of growth of current, on this hypothesis, from the rate of plasma ionization to be expected if ionization takes place by direct electron collision. In general, discussion is restricted to times of the order of 10^{-7} sec., for vapour from the cathode and positive ions moves a negligible distance in this time. Recombination of ions is also neglected. If it is also assumed that the ionized gas between anode and cathode is a genuine "plasma", i.e. that it consists of equal densities of positive ions and electrons, the calculation is simple, for then space charge effects may be neglected. Potential measurements by means of probes placed along the tube indicate that at any instant the plasma field is approximately equal to the tube potential difference divided by the electrode spacing (table 1).

If we then assume that in the plasma ionization initially takes place only by direct electron collision, the equations of continuity are (Thomson 1933)

$$\frac{\partial n_-}{\partial t} + \frac{\partial}{\partial x}(n_- u_-) = \alpha n_- u_-, \quad \dots\dots(6)$$

$$\frac{\partial n_+}{\partial t} - \frac{\partial}{\partial x}(n_+ u_+) = \alpha n_- u_-, \quad \dots\dots(7)$$

n_- , n_+ = density of electrons, positive ions,

u_- , u_+ = drift velocity of electrons, positive ions,

α = ionization efficiency (i.e. the number of ion pairs produced per electron per cm. of advance).

x refers to distance from the cathode.

$$\text{If } n_+ = n_-, \quad \frac{\partial}{\partial x}(n_- u_-) = - \frac{\partial}{\partial x}(n_+ u_+) = 0, \quad \dots\dots(8)$$

$$\text{and} \quad \frac{\partial n_-}{\partial t} = \alpha n_- u_-, \quad \dots\dots(9)$$

or, if the discharge path is not of unit cross-section, $\partial N/\partial t = \alpha N U$, where U now refers to electron drift velocity, and N is the actual number of ions present per unit length. Hence,

$$N = N_0 \exp\left(\int_0^t \alpha U dt\right),$$

where N_0 is the initial number of ions present. But $NqU = i$, where i = current and q = electron charge, whence

$$i = N_0 q U \exp\left(\int_0^t \alpha U dt\right). \quad \dots\dots(10)$$

Table 1

	Helium			Argon		
Gas pressure (mm. Hg)	15	8	0.5	8	3	0.5
Initial condenser voltage	310	310	310	310	310	310
X_p (v/cm.)	39.2	37.0	23.1	46.1	41.5	34.6
d (cm.)	3.0	3.0	3.0	3.0	3.0	3.0
$X_p d$	117	111	69.3	138	125	104
Tube p.d. at $(di/dt)_{\max}$ (v.)	155	155	108	193	180	135

X_p is the gradient in the plasma at $(di/dt)_{\max}$; d is the electrode spacing.

NOTE.— The plasma potential gradient is obtained from the main differences in potential indicated by four identical "floating" probes equally spaced along the body of the tube. The tube used had a mercury cathode. The circuit impedances were slightly different from those used in the measurements of figures 3, 4, 5.

In conditions of no space charge the electric field and current in the tube at any time are independent of the position in the tube, i.e. X , U , and α are independent of x .

If in equation (10) we measure time from the moment when $i = 1$ amp., we get

$$1 = N_0 U_0, \quad \dots\dots(11)$$

whence

$$i = (U/U_0) \exp\left(\int_0^t \alpha U dt\right),$$

and $U = KX$ where K is the electron mobility in an electric field X . The equation we have to solve is thus

$$i = (X/X_0) \exp\left(\int_0^t \alpha U dt\right). \quad \dots\dots(12)$$

Unfortunately it cannot be assumed that X , and hence α , are constant with time since the inductive drop of voltage in the leads and the resistive drop cause X to change rapidly with time.

If the capacity of the condenser is assumed to be large, so that a negligible charge flows out of it during the early stages of the discharge in which we are most interested, we can put

$$X = d^{-1}(V_0 - 2V_i - L(di/dt) - iR), \quad \dots\dots(13)$$

where d is the electrode spacing, V_0 the initial applied voltage, V_i the ionizing potential of the gas (assumed cathode and anode fall), L the total inductance and R the total resistance of the leads.

α also varies with X/P where P is the residual gas pressure. For various values of X/P , α has been fairly accurately evaluated by several workers (Druyvesteyn and Penning 1940), so that equation (12) can be solved by a laborious numerical integration. In obtaining numerical results one is hampered by uncertainty in the values for the electron mobility K found by other workers. It appears to have been measured only at values of X/P lower than those used here, and although it is fairly constant over the measured range, some error may be introduced by extrapolation; this is, however, not likely to be great enough to alter seriously the conclusions reached.

In figures 3-5 it is seen that the current rises at rates expected on the above hypotheses over a limited pressure range. At higher pressures (e.g. 5.5 mm. air) the measured rises are more rapid than expected: some additional form of ionization must be operative, the most probable being some "thermal" mechanism. This appears to have appreciable effect at values of X/P lower than about 10 for helium, 30 for argon, and 40 for air. Ionization by direct collision will account for the measured current growths for values of X/P probably up to at least 150 for these gas fillings. It must be emphasized that these results are obtained for pressure ranges of 2 to 40 mm. Hg for helium, 2 to 6 mm. for air, and 2 to 30 mm. for argon. A substantial triggering pulse is needed at the higher pressure end of these measurements, but once the arc has started, this pulse has never been found to have effect upon the current-time characteristics of the resulting discharge. Within these limits the cathode spot appears to have no effect.

ACKNOWLEDGMENTS

The writer is indebted to Prof. Sir George P. Thomson, F.R.S., for many valuable discussions, to Dr. M. Blackman for suggestions on the heating of the cathode, to Mrs. K. D. Froome for assistance in the preliminary experimental work and preparation of diagrams, and to B. A. Jarrett for considerable advice.

REFERENCES

- DRUYVESTEYN, M. J., and PENNING, F. M., 1940, *Rev. Mod. Phys.*, **12**, 88.
FROOME, K. D., 1946, *Nature, Lond.*, **157**, 446; 1947, *Ibid.*, **159**, 129.
FROOME, K. D., and JARRETT, B. A., 1943, *J. Roy. Photog. Soc.*, **83**, 352; 1944, *J. Roy. Microsc. Soc.*, **64**, 136.
HENRY, P. S. II., 1944, *J. Sci. Instrum.*, **21**, 135.
LANGMUIR, I., 1943, *Gen. Elect. Rev.*, **26**, 731.
SLEPIAN, J., 1942, *J. Appl. Phys.*, **13**, 113.
THOMSON, J. J. and G. P., 1933, *Conduction of Electricity through Gases* (Cambridge: University Press), Vol. 2, p. 512.

Blue and Ultra-Violet Bands of K_2

By S. P. SINHA

Imperial College, London

MS. received 9 July 1947

ABSTRACT. Bands of K_2 have been photographed in absorption in the region λ 4510– λ 3940 Å. in the first order of a 21-ft. concave grating with a dispersion of about 1.28 Å. per mm. The bands measured fall into two systems, one lying between λ 4510 and λ 4220 Å. and the other between λ 4160 and λ 3940 Å. They can be represented by the following equations :

$$\begin{aligned}\lambda\ 4510\text{--}\lambda\ 4220\ \text{\AA.} : \nu &= 22970.0 + 60.60(u') - 0.20(u'^2) - 92.64(u'') + 0.354(u''^2), \\ \lambda\ 4160\text{--}\lambda\ 3940\ \text{\AA.} : \nu &= 24627.7 + 61.60(u') - 0.90(u'^2) + 0.0010(u'^3) - 0.00030(u'^4) \\ &\quad - 92.64(u'') + 0.354(u''^2),\end{aligned}$$

where $u = v + \frac{1}{2}$.

The upper states of the two systems are considered to dissociate into $4\ ^2S + 5\ ^2P$ and $4\ ^2S + 3\ ^2D$ atoms respectively.

§1. INTRODUCTION

POTASSIUM is known to possess a number of systems of bands in the infra-red, visible and the ultra-violet region. The infra-red bands were first photographed by McLennan and Ainslie (1923) and their vibrational analysis has been carried out by Ritschl and Villars (1928) and by Crane and Christy (1930). The vibrational structure of the red system has been studied by Fredrickson and Watson (1927), and by Crane and Christy (1930) in absorption and by Loomis and Nusbaum (1932) who used the technique of magnetic rotation, while the rotational structure of this system has been studied by Loomis (1931). A vibrational analysis of the blue bands has been made by Weizel and Kulp (1930) from the data of Walter and Barratt (1928) and by Yamamoto (1929). Several systems of bands have been reported in the ultra-violet region by Chakraborti (1936), Yoshinaga (1937) and Sinha (1945, 1945 a). In general the analyses of the ultra-violet bands provided by the various authors do not appear satisfactory in as much as the systems look fragmentary. A somewhat similar situation existed in case of the ultra-violet bands of Na_2 , but it has now been shown that the bands which were analysed into five different systems by Weizel and Kulp (1930) constitute in reality two systems only (Sinha 1947). The present investigation has therefore been undertaken with a view to improving the existing data and obtaining a more satisfactory analysis of the bands of K_2 in the blue and ultra-violet region. The work has so far been carried out only in the region λ 4500– λ 3900 Å., although bands have been reported in previous works to wavelengths as low as λ 2900 Å. Work is in progress, however, in the latter region.

§2. EXPERIMENTAL

The bands have been studied in absorption. The absorption vessel consisted of a cylindrical steel tube, 65 cm. long and 2.5 cm. in internal diameter, provided with water-cooled glass or quartz windows. The central portion of the tube (30 cm. in length) could be electrically heated from outside to temperatures up to 1000° C. Potassium, freed from the oil in which it was stored by scraping off the outer layer of the oxide and washing two to three times in ether containing a few per cent. of alcohol, was put into a steel tube which was

then pushed to the centre of the absorption tube. The absorption tube was evacuated with a Cenco hyvac pump and filled with nitrogen at a pressure of a few cm. of mercury. The presence of nitrogen prevents rapid distillation of the metal to cooler parts when the tube is heated. A pointolite lamp was used to give continuous radiation. When used in proper orientation, the continuum given by it was free from band structure.

Preliminary investigations to determine the conditions under which the bands appeared most satisfactorily in different regions were first investigated with a low dispersion spectrograph. It was found that with a given amount of nitrogen gas inside the absorption tube, a higher temperature was generally needed to produce the systems the smaller the wavelength. Thus when the total pressure of nitrogen and potassium vapour in the tube, as indicated by a manometer connected to it was 10–15 cm. Hg, the blue bands appeared satisfactorily at about 550° C. (within the system a slightly higher temperature was needed to obtain the long wavelength end and a slightly lower temperature to obtain the short wavelength end satisfactorily), and the bands between λ 4160 and λ 3940 Å. at about 700° C. Using these values of temperature and pressure, the spectrum was next photographed in the first order of a 21-ft. concave grating (Eagle mounting), having a dispersion of about 1.28 Å. per mm. The iron arc spectrum has been used for comparison.

§3. APPEARANCE OF THE BANDS AND MEASUREMENTS

The bands appear in two distinct regions: (i) λ 4510– λ 4220 Å. and (ii) λ 4160– λ 3940 Å. They are all degraded to the red. The heads appear much better

Table 1. K_2 bands: λ 4510– λ 4220 Å.

λ_{air} (Å.)	Int.	ν', ν''	$\nu_{obs} - \nu_{calc}$ (cm^{-1})	λ_{air} (Å.)	Int.	ν', ν''	$\nu_{obs} - \nu_{calc}$ (cm^{-1})
4505.3	1	2,10	1	4343.5	10	1,0	0
4500.0	2	1,9	3	4338.1	5	3,1	2
4495.1	3	0,8	0	4332.3	7	2,0	0
4487.8	2	2,9	3	4327.1	4	4,1	2
4482.8	2	1,8	1	4320.9	7	3,0	2
4477.7	4	0,7	0	4316.2	4	5,1	1
4470.6	3	2,8	2	4310.0	7	4,0	0
4465.6	3	1,7	-1	4304.8	4	6,1	3
4460.3	5	0,6	-1	4299.0	8	5,0	1
4453.7	2	2,7	0	4294.5	5	7,1	2
4448.2	3	1,6	0	4288.4	8	6,0	1
4442.6	6	0,5	0	4284.0	4	8,1	2
4436.5	4	2,6	1	4277.6	6	7,0	1
4430.8	4	1,5	1	4273.6	6	9,1	1
4425.5	6	0,4	-2	4269.8	3	11,2	1
4419.3	4	2,5	-1	4267.2	4	8,0	1
4413.3	3	1,4	0	4263.6	4	10,1	0
4407.7	5	0,3	-1	4259.5	2	12,2	1
4401.8	4	2,4	0	4256.8	3	9,0	2
4395.9	4	1,3	0	4253.0	2	11,1	2
4390.2	5	0,2	-1	4249.6	3	13,2	1
4378.4	6	1,2	0	4243.4	3	12,1	-1
4372.9	7	0,1	-2	4236.7	2	11,0	1
4367.8	5	2,2	1	4229.8	1	15,2	1
4361.0	6	1,1	0	4223.0	1	14,1	3
4355.1	8	0,0	-1				
4349.7	6	2,1	0				

marked in the former region than in the latter. Lines due to rotational structure appear over the entire region, but the branches cannot be identified at all owing to overlapping and closeness of structure.

Measurements of the heads of the bands are given in tables 1 and 2. Intensities given are visual estimates on a scale of 1–10; they have been estimated in different regions (separated by horizontal lines in the tables) from different spectrograms.

Table 2. K_2 bands: λ 4165– λ 3940 Å.

λ_{air} (Å.)	Int.	ν', ν''	$\nu_{\text{obs}} - \nu_{\text{calc}}$ (cm^{-1})	λ_{air} (Å.)	Int.	ν', ν''	$\nu_{\text{obs}} - \nu_{\text{calc}}$ (cm^{-1})
4164.7	1	5,10	–1	4063.5	4	3,2	–1
4161.0	1	—	—	4059.9	4	5,3	3
4159.1	1	4,9	–1	4059.0	4	7,4	–2
4155.4	2	6,10	0	4057.9	5	2,1	–2
4153.2	2	3,8	–1	4054.8	4	4,2	–3
4149.4	2	5,9	–2	4033.5	6	3,0	–1
4146.0	3	7,10	–4	4031.1	4	5,1	–2
4144.3	3	4,8	–2	4029.3	2	—	—
4138.2	3	3,7	0	4024.9	6	4,0	–3
4137.5	3	8,10	4	4022.8	4	6,1	–2
4134.6	4	5,8	1	4016.3	5	5,0	–2
4133.1	4	2,6	0	4014.4	5	7,1	0
4131.1	4	7,9	6	4008.0	5	6,0	–2
4127.6	5	1,5	1	4004.9	3	—	—
4125.6	4	6,8	2	3999.6	5	7,0	0
4123.6	5	3,6	–1	3995.1	4	14,3	–3
4122.7	6	0,4	1	3992.6	4	12,2	–1
4121.9	4	—	—	3991.6	5	8,0	1
4119.8	5	5,7	0	—	—	10,1	1
4115.1	3	—	—	3989.6	3	15,3	–4
4113.8	4	4,6	3	3986.2	3	13,2	2
4112.8	8	1,4	0	3984.7	5	11,1	1
4110.5	4	6,7	4	3984.1	4	9,0	2
4108.6	6	3,5	–1	3980.8	4	—	—
4107.3	7	0,3	2	3979.8	4	14,2	2
4104.7	4	5,6	3	3978.9	4	—	—
4103.0	6	2,4	0	3978.0	5	12,1	0
4101.7	4	—	—	3976.7	4	10,0	–3
4099.1	5	4,5	0	3975.8	3	—	—
4097.4	7	1,3	1	3974.3	3	15,5	0
—	—	6,6	–3	—	—	—	—
4095.7	4	—	—	3972.1	3	13,1	1
4093.9	5	3,4	–2	3970.3	3	11,0	0
4092.3	8	0,2	–2	3969.6	3	16,2	–6
4090.2	4	5,5	0	3966.0	3	14,1	–2
4087.5	6	2,3	2	3964.6	3	12,0	–6
4084.7	4	4,4	–2	3963.2	2	17,2	0
4082.7	10	1,2	–1	3960.6	3	15,1	0
4081.4	5	6,5	2	3957.2	3	13,0	1
4078.2	6	3,3	1	3955.3	2	16,1	–7
4075.5	5	5,4	0	3951.6	1	14,0	–2
4073.0	4	2,2	–1	3946.3	1	15,0	–5
4069.1	2	4,3	2	3941.3	1	—	—
4067.0	8	1,1	1	—	—	—	—
—	—	6,4	–1	—	—	—	—
4065.1	3	—	—	—	—	—	—

§4. VIBRATIONAL ANALYSIS AND MOLECULAR CONSTANTS

All the bands measured fall into two systems, one lying between λ 4510 and λ 4220 Å. and the other between λ 4160 and λ 3940 Å. We will, for convenience,

call them the blue and the first ultra-violet systems respectively. The assignments of vibrational quantum numbers are given in tables 1 and 2 and the arrangements in v' , v'' schemes in tables 3 and 4 respectively.

Table 3. Deslandres' scheme for the blue system of K_2 bands

v''	0	1	2	3	4	5	6	7	8	9	10	Mean diff.
8	22955 61	22862 62	22772 61	22681 61	22590 62	22503 60	22414 61	22327 60	22240 60			60.9
	23016 60	22924 60	22833 61	22742 61	22652 60	22563 60	22475 59	22387 60	22301 61	22216 60		60.0
	23076 61	22984 59	22894 61		22712 60	22622 59	22534 61	22447 60	22362 61	22276 60	22190 98	60.0
	23137 58	23045 59										58.5
	23195 59	23104 58										58.5
	23254 58	23162 61										59.5
	23312 58	23223 56										57.0
	23370 58	23279 57										57.5
	23428 57	23336 57										57.0
	23485 (57)	23393 55										56.0
	—	23448 58										56.5
1	23597 (55)	23506 53	23414 56									54.5
		23559 57	23470 55									56.0
		23616 57	23525 (55)									56.0
		23673 (55)	—									55.0
			23635 (55)									
an diff.	91.6	90.5	91.0	90.5	88.7	88.3	87.3	86.0	85.5	86.0		

Table 4. Deslandres' scheme for the first ultra-violet system of K₂ bands

v''	0	1	2	3	4	5	6	7	8	9	10	Mean diff.
0			24429	24340	24249							
1	*	24581	24487	24399	24308	24220						58.7
2	*	24636	24545	24458	24366		24188					57.5
3		*	24785	24602	24420	24332	24244	24158	24071			55.3
4		*	24838	24655	24475	24389	24302		24123	24037		54.9
5		24800	24655	24568	24475	24389	24302		24123	24037		54.4
6		24943	24851	24662	24530	24442	24355	24266	24179	24093	24005	52.6
7		24903	24800	24622	24530	24442	24355	24266	24179	24093	24005	51.7
8		25045	24943	24751	24630	24512	24409	24321	24232	24142	24058	50.0
9		25093	24995	24895	24795	24695	24595	24495	24395	24295	24195	48.0
10		25139	25045	24945	24845	24745	24645	24545	24445	24345	24245	46.0
11		25180	25089	24989	24889	24789	24689	24589	24489	24389	24289	42.5
12		25225	25131	25039	24942	24842	24742	24642	24542	24442	24342	43.5
13		25263	25169	25079	24983	24883	24783	24683	24583	24483	24383	38.7
14		25299	25207	25107	25007	24907	24807	24707	24607	24507	24407	37.0
15		25333	25242	25142	25042	24942	24842	24742	24642	24542	24442	35.3
16		25365	25275	25184	25093	24993	24893	24793	24693	24593	24493	32.3
17												32.0
Mean diff.	92.2	91.6	88.7	91.8	87.2	87.0	87.7	87.6	86.0	86.0		

(a) *Blue system*

As mentioned before, the blue bands were analysed by Yamamoto (1929) and by Weizel and Kulp (1930). The analysis of the former appears more satisfactory and is in good agreement with the present one.

All the blue bands measured in the present case can be represented by the equation

$$\nu = 22970.0 + 60.60u' - 0.20u'^2 - 92.64u'' + 0.354u''^2 \quad \dots\dots(1)$$

where $u = v + \frac{1}{2}$.

The terms in u'' are those from the work of Loomis and Nussbaum (1932). The values of $\nu_{\text{obs}} - \nu_{\text{calc}}$ given in table 1 indicate fairly good agreement between the observed positions of the band heads and those calculated from equation (1).

The heat of dissociation for the upper state calculated from the above formula by applying Birge and Sponer's extrapolation rule is 4560 cm^{-1} . This enables us to calculate ν_{atom} , the energy of the dissociated atoms, given by the expression, $\nu_{\text{atom}} = \nu_{0,0} + D' - D''$, where the symbols used have the usual significance. Now $\nu_{0,0} = 22955 \text{ cm}^{-1}$, $D' = 4560 \text{ cm}^{-1}$ and $D'' = 4130 \text{ cm}^{-1}$, and therefore $\nu_{\text{atom}} = 23385 \text{ cm}^{-1}$. Further, from the study of line spectra data for potassium, we have $4^2S - 5^2S = 21025 \text{ cm}^{-1}$, $4^2S - 3^2D = 21540 \text{ cm}^{-1}$ and $4^2S - 5^2P = 24710 \text{ cm}^{-1}$. Comparison of these with the above value of ν_{atom} shows that the upper state of the blue system dissociates into 4^2S and 5^2P atoms of K.

(b) The first ultra-violet system

Bands in this region have previously been reported by Yoshinaga, who regards them as belonging to two different systems, one lying in the region $\lambda 4160 - \lambda 4050 \text{ \AA.}$ and the other in the region $\lambda 4040 - \lambda 3960 \text{ \AA.}$; his systems are fragmentary and comparison of his measurements with table 4 indicates that all the bands which he measured in this region could be included in a single system. It appears that the presence of the $\lambda 4045.6 \text{ \AA.}$ line of the P series of K which obscures some bands between $\lambda 4040$ and $\lambda 4050 \text{ \AA.}$ led him to consider the bands as belonging to two different systems. The region of overlap of the bands by the broadened P series line is indicated by asterisks in table 4.

The bands measured in the present investigation can be represented by the following equation:

$$\nu = 24627.7 + 61.60u' - 0.90u'^2 + 0.0010u'^3 - 0.00030u'^4 - 92.64u'' + 0.354u''^2, \quad \dots\dots(2)$$

where $u = v + \frac{1}{2}$ as before.

The terms in u'' have been retained, as before, from the work of Loomis and Nussbaum. Values of $\nu_{\text{obs}} - \nu_{\text{calc}}$ given in table 2 indicate fairly good agreement between the observed frequencies and those calculated from equation (2).

The value of D' cannot in this case be obtained from Birge and Sponer's extrapolation rule. It has therefore been determined from figure 1, which gives a plot of T' (i.e. $\nu_{0,0}, \nu_{1,0}, \nu_{2,0}, \dots$) against $\Delta T'$ (i.e. $\nu_{0,0} - \nu_{1,0}, \nu_{1,0} - \nu_{2,0}, \dots$). T'_{max} , the value of T' corresponding to $\Delta T' = 0$, is equal to 25590 cm^{-1} , and since $\nu = 24612 \text{ cm}^{-1}$, D' is approximately 980 cm^{-1} .

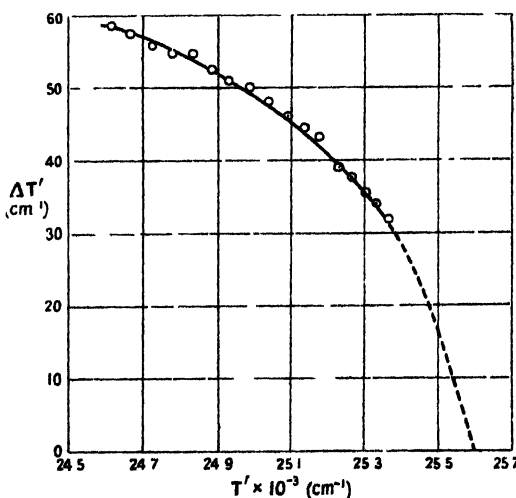


Figure 1.

$\Delta T - T'$ curve for $\lambda 4160 - \lambda 3940$ bands of K_2 .

The energy of the dissociated atoms can be determined from the relation, $\nu_{\text{atom}} = T'_{\text{max}} - D'$ and is equal to 21455 cm^{-1} . Comparing this value of ν_{atom} with the separations of the excited atomic K levels from its ground 4^2S level, given above, it may be concluded that the dissociation products for the excited state of this system $1 K_2$ bands are 4^2S and 3^2D atoms of K.

Vibrational constants for the five states of K_2 are summarized in table 5. The constants ν_e , ω_e , $x_e\omega_e$, $y_e\omega_e$, $z_e\omega_e$ are those in the expression (3) and are given in cm^{-1}

$$\nu = \nu_e + \omega_e v - x_e \omega_e v^2 + y_e \omega_e v^3 - z_e \omega_e v^4, \quad (3)$$

where $v = v' + \frac{1}{2}$.

Table 5. Vibrational constants for the ground and the excited states of K_2

States	ν_e	ω_e	$x_e\omega_e$	$y_e\omega_e$	$z_e\omega_e$	D	Dissociation products
$1\Sigma_g^+$		92.64	0.354	—	—	4135	$4^2S + 4^2S$
$1\Sigma_u^+$	11683.6	69.09	0.153	—	—	5510	$4^2S + 4^2P$
$1\Pi_u$	15378.0	75.00	0.388	0.00437	0.00018	1780	$4^2S + 4^2P$
$1\Sigma_u^-$ (?)	22970.0	60.60	0.20	—	—	4560	$4^2S + 5^2P$
	24627.7	61.60	0.90	0.0010	0.00030	980	$4^2S + 3^2D$

§5. CONCLUSION

The ground state of K_2 is $1\Sigma_g^+$, due to two atoms of potassium in the normal 4^2S state. The next two states are $1\Sigma_u^+$ and $1\Pi_u$ states formed from 4^2S and 4^2P atoms. Transitions between these and the ground state are responsible for the infra-red and the red systems of bands of K_2 . The next higher molecular states revealed in absorption experiment are those involved in the blue and the ultra-violet systems studied in the present investigation. The upper state of the blue system has been shown to be due to $4^2S + 5^2P$ atoms and from analogy with the states due to $4^2S + 4^2P$ atoms, it may be concluded that this state is of $1\Sigma_u^+$ type. The blue system should thus be due to $1\Sigma_u^+ \leftarrow 1\Sigma_g^+$ transition and the bands should be single headed and should only have two branches. The general appearance of the bands suggests that they are single headed, but confirmation is impossible since the dispersion used was insufficient to enable resolution of two separate heads. An attempt was made to employ sufficient dispersion so as to see clearly the branches possessed by the bands. This has not, however, succeeded. The upper state of the next system, i.e. the first ultra-violet system, is regarded as due to $4^2S + 3^2D$ atoms of potassium. It is, however, also equally probable that it may be due to $4^2S + 5^2S$ atoms. At wavelengths lower than $\lambda 3900 \text{ \AA}$, several systems of bands have been proposed by different workers (referred to earlier), but only one system appears extensive ($\lambda 3900 - \lambda 3700 \text{ \AA}$). It is not, however, possible to make a correct estimate of D' even for this system, and thereby to determine the states of the atoms into which its upper state dissociates. If however, we suppose that this state dissociates into $4^2S + 5^2P$ atoms, the value of D' required for this purpose would be about 2000 cm^{-1} , and the state would be $1\Pi_u$ in character. A comparison with the value of dissociation energy of $1\Pi_u$ state from $4^2S + 4^2P$ atoms given in table 5 indicates that the above value of D' is not unlikely. Investigations are in progress to clarify the position with regard to the bands of this system and of other systems at lower wavelengths.

ACKNOWLEDGMENTS

Finally, it is a great pleasure to express my gratitude to Assistant Professor R. W. B. Pearse, D.Sc., for his valuable guidance in the work. My thanks are also due to the University of Patna, India, for the award of their Birla scholarship, which has enabled me to carry out the work.

REFERENCES

- CHAKRABORTY, B. K., 1936, *Ind. J. Phys.*, **10**, 155.
 CRANE, W. O., and CHRISTY, A., 1930, *Phys. Rev.*, **36**, 421.
 FREDRICKSON, W. R., and WATSON, W. W., 1927, *Phys. Rev.*, **30**, 428.
 LOOMIS, F. W., 1931, *Phys. Rev.*, **38**, 2152.
 LOOMIS, F. W., and NUSBAUM, R. E., 1932, *Phys. Rev.*, **39**, 89.
 MCLENNAN, J. C., and AINSLIE, D. S., 1923, *Proc. Roy. Soc. A*, **103**, 304.
 RITSCHL, R., and VILLARS, D., 1928, *Naturwissenschaften*, **16**, 219.
 SINHA, S. P., 1945 a, *Curr. Sci.*, **14**, 230 ; 1945 b, *Thesis for Ph.D.*, Patna Univ., India ;
 1947, *Proc. Phys. Soc.*, **59**, 610.
 WALTER, J. M., and BARRATT S., 1928, *Proc. Roy. Soc. A*, **119**, 257.
 WEIZEL, W., and KULP, M., 1930, *Ann. Phys., Lpz.*, **4**, 971.
 YAMAMOTO, H., 1929, *Jap. J. Phys.*, **5**, 153.
 YOSHINAGA, H., 1937, *Proc. Phys. Math. Soc., Japan*, **19**, 847.

Ultra-Violet Bands of Li_2

By S. P. SINHA

Imperial College, London

MS. received 1 August 1947

ABSTRACT. The ultra-violet bands of Li_2 have been photographed in absorption on a quartz spectrograph, using lithium vapour in argon at a pressure of half an atmosphere, at temperatures between 900 and 1100° c. Bands appear between λ 3500 and λ 3100 Å. The band heads, which have been measured between λ 3450 and λ 3130 Å., can be represented by the equation :

$$\nu = 30658.5 + 231.5u' - 1.5u'^2 - 351.6u'' + 2.6u''^2.$$

where $u = v + \frac{1}{2}$.

The upper state of this system of bands is considered to dissociate into $2^2S + 3^2P$ atoms of lithium.

§1. INTRODUCTION

THERE are two systems of Li_2 bands in the visible region : (i) λ 7700– λ 6550 Å. and (ii) λ 5600– λ 4500 Å. These have been studied by Wurm (1929), Harvey and Jenkins (1930), Loomis and Nusbaum (1931) and others, and are due to $^1\Sigma_u^+ \leftarrow ^1\Sigma_g^+$ and $^1\Pi_u \leftarrow ^1\Sigma_g^+$ transitions. Li_2 has also some bands in the ultra-violet region between λ 3500 and λ 3100 Å. which have been considered to belong to four different systems (Vance and Huffman 1935). These systems, however, appear fragmentary. Moreover, the corresponding bands of Na_2 —the λ 3600– λ 3200 Å. system—formerly regarded as due to several systems of bands have now been shown to constitute a single system (Sinha 1947). It has therefore been considered necessary to re-photograph the lithium ultra-violet bands and attempt a fresh analysis.

§2. EXPERIMENTAL

The details of the absorption tube used in the present investigation have been described earlier in connection with the ultra-violet bands of K_2 (Sinha 1948). Lithium, freed from the oil in which it was stored, was placed at the centre of the tube and heated electrically from outside to temperatures up to 1100°C . The source of continuous radiation was a 6-v. 18-amp. tungsten lamp which gave a fairly intense continuum down to $\lambda 3100\text{ \AA}$. A quartz spectrograph (E_1) was used to photograph the bands.

In the case of sodium and potassium, experience had shown that the presence of a certain amount of a gas like nitrogen, which did not react chemically with the metal, was almost a necessity to obtain a satisfactory photograph of the ultra-violet bands, and nitrogen was invariably used. The main function of the gas is probably to prevent a rapid distillation of the metal to cooler parts of the absorption tube. With lithium, however, it was found that as soon as the metal was heated, all nitrogen introduced inside the chamber was absorbed by it and the pressure was reduced to its initial value. Under these conditions, although the visible bands appeared, there was no more than a trace of the ultra-violet bands, nor was there any spectroscopic evidence of the formation of a lithium-nitrogen compound. Nitrogen had, therefore, to be discarded and argon used. This was not absorbed by hot lithium and in its presence the Li_2 bands appeared quite satisfactorily in the ultra-violet region. When the pressure, as read on a manometer connected to the absorption chamber, was about 40 cm. Hg, the bands appeared well developed at temperatures between 900 and 1100°C . (estimated from the colour of the furnace).

§3. APPEARANCE OF THE SPECTRUM AND MEASUREMENTS

The bands, which appear between $\lambda 3500$ and $\lambda 3100\text{ \AA}$., are all degraded to the red. Lines due to rotational structure can be seen throughout the region, but it is not possible to draw any conclusion with regard to the number of branches they have, unless the bands can be photographed at a higher dispersion without much overlapping. It is equally difficult to say whether they are single-headed; for even if they were double-headed, the heads would lie too close together to be seen distinctly resolved at the dispersion employed. The iron arc has been used for comparison. Wavelengths of the band heads are given in table 1; intensities given are visual estimates on a scale of 1-10.

§4. VIBRATIONAL ANALYSIS AND MOLECULAR CONSTANTS

All the band-heads measured are found to belong to a single system. The vibrational quantum numbers assigned to them are given in table 1, and table 2 gives their arrangement in a Deslandres scheme. The heads can be represented by the equation:

$$\nu = 30658.5 + 231.5(u') - 1.5(u'^2) - 351.6(u'') + 2.6(u''^2),$$

where $u = v + \frac{1}{2}$.

The terms in u'' are those of Loomis and Nusbaum (1931). Values of $\nu_{\text{obs}} - \nu_{\text{calc}}$ given in table 1 indicate that the representation of the band heads by the above equation is satisfactory.

The value of the energy of dissociation for the upper state, calculated from Birge and Spomer's rule, is about 9000 cm^{-1} ; the value thus calculated can be

Table 1. Li_2 ultra-violet bands

λ_{air} (Å.)	Int.	v', v''	$\nu_{\text{obs}} - \nu_{\text{calc}}$	λ_{air} (Å.)	Int.	v', v''	$\nu_{\text{obs}} - \nu_{\text{calc}}$
3442.8	2	1,5	1	3266.3	2	2,1	0
3431.2	4	0,4	-3	3253.1	10	1,0	3
3416.3	1	2,5	0	3229.6	2	2,0	2
3404.4	4	1,4	-3	3219.9	1	4,1	0
3392.1	6	0,3	-2	3206.9	6	3,0	-1
3378.5	4	2,4	-3	3197.2	1	5,1	3
3365.8	3	1,3	-1	3184.2	6	4,0	1
3353.6	10	0,2	-1	3175.3	2	6,1	4
3340.5	2	2,3	-2	3162.0	5	5,0	3
3327.8	2	1,2	1	3154.5	2	7,1	3
3315.6	9	0,1	0	3140.8	4	6,0	-1
3290.5	4	1,1	1	3133.8	2	8,1	1
3277.6	6	0,0	2				

Table 2. Arrangement of Li_2 ultra-violet bands in Deslandres' scheme

v'	v''	0	1	2	3	4	5	Mean diff.
0		30501	30152	29810	29472	29136		
		230	230	231	230	229		230
1		30731	30382	30041	29702	29365	29038	
		224	225		225	225	225	224.8
2		30955	30607		29927	29590	29263	
		219				337	327	219
3		31174						
		222						222
4		31396	31048					
		220	220					220
5		31616	31268					
		214	213					213.5
6		31830	31481					
			211					211
7			31692					
			208					208
8			31900					
Mean diff.		348.5	341.5	338.5	336.7	327		

accepted only with great reserve, since the calculation involves a large range of extrapolation.

§5. ISOTOPE EFFECT

Lithium has two isotopes, ^6Li and ^7Li and their relative abundance is in the approximate ratio 1 to 12. It would therefore appear that although it may not be possible to see bands due to $^6\text{Li}_2$, strong bands of $^6\text{Li}^7\text{Li}$ may possibly be seen under favourable conditions. The separation between the heads of bands due to $^6\text{Li}^7\text{Li}$ and $^7\text{Li}_2$ may theoretically be calculated from the expression

$$\nu' - \nu = (\rho - 1)[\omega_v'(u') - \omega_v''(u'')],$$

where $\rho = \sqrt{(\mu/\mu')} = 1.041$ approximately, and the superscript i refers to the $^6\text{Li}^7\text{Li}$ molecule. Calculation from the above expression indicates that, of the bands of $^7\text{Li}_2$ molecule which have appeared strongly on the plate, the 0,4 0,3 0,2 and 0,1 bands will have the components due to $^6\text{Li}^7\text{Li}$ on their long-wave side, i.e. on the side on which their own rotational fine structure will extend, and it will therefore be difficult to detect them. For the remaining strong bands, i.e. the 1,0 3,0 and 4,0 bands, the isotope components will lie on the short-wave side and may possibly be detected. Faint heads separated from the strong

heads of 3,0 and 4,0 bands by 24 and 35 cm^{-1} have been detected on the plate, the corresponding calculated values being 25 and 34 cm^{-1} respectively.

§6. DISSOCIATION PRODUCTS

The energy of the dissociated atoms from the excited state of the ultra-violet system can be calculated from the expression: $\nu_{\text{atom}} = \nu_{0,0} + D' - D''$, where the terms used have the usual significance. Since $\nu_{0,0} = 30500 \text{ cm}^{-1}$, $D' = 9000 \text{ cm}^{-1}$ and $D'' = 9200 \text{ cm}^{-1}$ (Loomis and Nusbaum 1931), we have $\nu_{\text{atom}} = 30300 \text{ cm}^{-1}$. Also from Li line spectrum data, we have $2^2S - 3^2S = 27205 \text{ cm}^{-1}$, $2^2S - 3^2P = 30926 \text{ cm}^{-1}$ and $2^2S - 3^2D = 31283 \text{ cm}^{-1}$. A comparison of these with the above value of ν_{atom} suggests that the excited state dissociates into $2^2S + 3^2P$ atoms.

§7. DISCUSSION

Altogether three systems of bands due to Li_2 are now known, two of which are in the visible and one in the ultra-violet. The visible systems have been definitely established to be due to transitions between the ground $^1\Sigma_g^+$ ($2^2S + 2^2S$ atoms) and the excited $^1\Sigma_u^+$ and $^1\Pi_u$ ($2^2S + 2^2P$ atoms) states. Theoretical considerations indicate that the next higher excited states to which transitions from the ground $^1\Sigma_g^+$ state are permitted will be the following: (i) $^1\Sigma_u^+$ state from $2^2S + 3^2S$ atoms, (ii) $^1\Sigma_u^+$ and $^1\Pi_u$ states from $2^2S + 3^2P$ atoms and (iii) $^1\Sigma_u^+$ and $^1\Pi_u$ states from $2^2S + 3^2D$ atoms. Since, from simple considerations of Heitler-London theory, all these states would be stable, we should observe five systems of bands in the region under present investigation. A somewhat similar discrepancy between the number of systems expected from considerations of Heitler-London theory and the number observed has been noted also in Na_2 and K_2 . For each of these molecules three systems of bands have been found, two of which are believed to be due to $^1\Sigma_u^+$, $^1\Pi_u(m^2S + (m+1)^2P) - ^1\Sigma_g^+(m^2S + m^2S)$ and one due to transition from the ground $^1\Sigma_g^+$ state to an excited state dissociating into $m^2S + m^2D$ atoms ($m=3$ for Na_2 and for K_2). Of these three systems, that towards the long-wave side, i.e. the $\lambda 3600$ – $\lambda 3200$ Å. system of Na_2 and the $\lambda 4500$ – $\lambda 4200$ Å. system of K_2 , appears more intensely developed than the other two, and also appears at a lower temperature than is needed for the others. From analogy with these, it is considered probable that the Li_2 $\lambda 3500$ – $\lambda 3100$ Å. system corresponds to the Na_2 $\lambda 3600$ – $\lambda 3200$ Å. or to the K_2 $\lambda 4500$ – $\lambda 4200$ Å. system, and its upper state dissociates into $2^2S + 3^2P$ atoms. The systems of Li_2 bands corresponding to the other two systems of Na_2 or K_2 bands will require higher temperatures for their development and cannot be observed with the present steel absorption tube.

A few additional band-heads have been measured by Vance and Huffman which have led them to consider that there are three more systems of bands in the region investigated. Present observations, however, provide no evidence of the existence of these bands. Since Vance and Huffman have not published the intensities of the bands which they measured, it is difficult to account for this discrepancy.

ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to Assistant Professor R. W. B. Pearse, D.Sc., for his kind help and valuable suggestions, and to Patna University, India, for the grant of their Birla scholarship.

REFERENCES

- HARVEY, A., and JENKINS, F. A., 1930, *Phys. Rev.*, **35**, 789.
 LOOMIS, F. W., and NUSBAUM, R. E., 1931, *Phys. Rev.*, **38**, 1447.
 SINHA, S. P., 1947, *Proc. Phys. Soc.*, **59**, 610 ; 1948, *Ibid.*, **60**, 416.
 VANCE, J. E., and HUFFMAN, J. R., 1935, *Phys. Rev.*, **47**, 215.
 WURM, K., 1929, *Z. Phys.*, **59**, 35.

Ultra-Violet Bands of NaK

By S. P. SINHA

Imperial College, London

MS. received 1 August 1947

ABSTRACT. Bands due to NaK, in the region λ 3820 to λ 4080 Å., have been photographed in absorption in the first order of a 21-ft. concave grating and their vibrational analysis carried out. The constants for this system are (in cm^{-1})

$$\nu_0 = 25\,201, \quad \omega_0 = 95\cdot85, \quad x_0\omega_0 = 0\cdot94, \quad D' = 2320$$

$$\omega_0'' = 123\cdot29, \quad x_0\omega_0'' = 0\cdot40, \quad D'' = 5025.$$

The system is regarded as arising from transitions from the ground $^1\Sigma^+$ state to an excited state dissociating into $\text{Na } 3^2S + \text{K } 3^2D$ atoms.

§1. INTRODUCTION

THERE are three systems of bands of NaK in the infra-red and the visible regions: (i) λ 9150– λ 7200 Å., (ii) λ 6150– λ 5600 Å., and (iii) λ 5300– λ 4700 Å. They have been studied by Ritschl and Villars (1928) and by Loomis and Arvin (1933, 34). Some bands in the ultra-violet region, extending from λ 4020 to λ 3540 Å., attributed to NaK have been reported by Walter and Barratt (1928). Two different analyses of the ultra-violet bands have been proposed; one by Uchida (1929) and the other by Weizel and Kulp (1930). Since the data given by Walter and Barratt were obtained on an instrument of rather low dispersion and the two analyses of these data are widely divergent, it has been considered desirable to re-photograph the bands with a higher dispersion and attempt a fresh analysis.

§2. EXPERIMENTAL

The bands have been photographed in absorption. The details of the apparatus used are the same as for the ultra-violet bands of K_2 (Sinha 1948). Preliminary investigations were made on a small-dispersion instrument to study the conditions under which the NaK bands appeared satisfactorily in different regions. The observation of these bands is more difficult than for Na_2 or K_2 because they may be masked by bands due to these molecules if they have very strong absorption bands in any particular region. The difficulty may be overcome to a certain extent by adjusting the relative proportion of sodium and potassium in the absorption chamber. For example, in attempting to photograph NaK bands in a region likely to be overlapped by Na_2 bands, only a small proportion of sodium should be used. Thus, in the green and the yellow regions, both Na_2 and NaK are known to give bands; the former had to be suppressed in order to obtain NaK bands only and the trial experiments showed

that this was very effectively done by photographing the absorption spectrum of a sample of potassium in which the only sodium was that present as impurity. The bands appeared very satisfactorily at a temperature of about 500 to 600° c. when the pressure inside the absorption chamber, obtained by introducing nitrogen, was about 5 cm. Hg. Although the D lines appeared quite strongly in absorption, no trace of Na_2 bands was found. To photograph the ultra-violet bands of NaK , since overlapping by K_2 bands was expected, only a small proportion of potassium was used in the mixture. The bands appeared at a temperature of about 650° c. when the pressure of nitrogen inside the tube was about 15 cm. Hg. The method would be more difficult in a region where both Na_2 and K_2 bands were present, but the relative intensities of the bands as a function of the relative concentrations of sodium and potassium could be studied as a means of deciding which belonged to Na_2 , which to K_2 and which to NaK .

The bands were then photographed on high-dispersion instruments. For the visible bands, a tungsten lamp (6 v., 18 amp.) was used as a source of continuum and photographs were taken on a glass Littrow spectrograph (dispersion about 8 Å. per mm. at $\lambda 6000$ and 4 Å. per mm. at $\lambda 4500$). The ultra-violet bands were photographed in the first order of a 21-ft. concave grating (dispersion 1.3 Å. per mm.) using a 500 c.p. pointolite lamp to give the continuum.

§3. APPEARANCE OF THE SPECTRUM AND MEASUREMENTS

(a) Visible systems

As noted above the visible systems have been studied by a number of workers, but, since no satisfactory photograph of these bands has been published nor any satisfactory description given, it is considered worth while to publish the spectrograms which might prove useful for purposes of identification.

The visible band-systems of NaK , taken on a glass Littrow spectrograph, are reproduced in the plate facing p. 452. For the yellow system (strip (a)) between $\lambda 6150$ and $\lambda 5600$, the temperature was 500° c. and the pressure (by introducing nitrogen) 5 cm. Hg; the bands are all degraded to the red, and the heads can be clearly detected. The absorption seems stronger between $\lambda 5800$ and $\lambda 5600$ than at longer wavelengths. The $\lambda 5300$ – $\lambda 4700$ system (strips (b) and (c)) appear at about 600° c.; the bands are all degraded to the red. Rotational structure can be seen over the entire region, but the dispersion employed is small for its study.

Measurements and assignments of vibrational quantum numbers are given in tables 1 and 2; they agree with those due to Loomis and Arvin (1934). The intensities given are visual estimates on a scale of 10.

Table 1. NaK bands between $\lambda 5650$ and $\lambda 5970$

λ_{air} (Å.)	Int.	v', v''	λ_{air} (Å.)	Int.	v', v''	λ_{air} (Å.)	Int.	v', v''
5640.8	4	14,0	5745.4	9	7,0	5935.1	3	0,1
5650.4	5	13,0	5763.4	7	6,0	5954.2	3	1,2
5665.3	6	12,0	5783.0 *	6	5,0	5973.0	2	2,3
5676.0	3	18,2	5802.6	7	4,0	5997.0	3	1,3
5680.8	8	11,0	5823.5	6	3,0	6022.1	2	0,3
5696.2	8	10,0	5845.2	7	2,0	6040.5	2	1,4
5711.7	9	9,0	5869.5 *	5	1,0	6066.2	3	0,4
5728.2	10	8,0	5912.0 *	4	1,1			

* Bands not measured by Loomis and Arvin (1934).

Table 2. NaK bands between $\lambda 4730$ and $\lambda 5260$

λ_{air} (A.)	Int.	v', v''	λ_{air} (A.)	Int.	v', v''	λ_{air} (A.)	Int.	v', v''
4702.7	4	16,0	4847.4	10	7,0	5032.4	4	2,3
4717.8	4	15,0	4866.0	8	6,0	5052.6	4	1,3
4733.1	4	14,0	4876.8	10	7,1	5063.5	3	2,4
4748.5	4	13,0	4894.5	8	6,1	5083.8	4	1,4
4764.1	5	12,0	4913.5	8	5,1	5105.4	3	0,4
4779.7	6	11,0	4932.2	8	4,1	5115.0	2	1,5
4796.2 *	7	10,0	4943.2	5	5,2	5136.3	3	0,5
4813.0	8	9,0	4962.4	7	4,2	5167.5	2	0,6
4819.3	7	12,2	4982.2	7	3,2	5198.9	1	0,7
4830.3	10	8,0	5001.8	6	2,2	5230.5	1	0,8
4836.5	8	11,2	5011.9	5	3,3	5261.8	1	0,9

* Band not measured by Loomis and Arvin (1934).

(b) *Ultra-violet system*

Photographs taken on a small quartz instrument indicate that at wavelengths lower than $\lambda 3800$ A. only bands of Na_2 occur, and no additional bands attributable to NaK can be detected. There is, however, a patch of what appears at this dispersion to be continuous absorption between $\lambda 3820$ and $\lambda 4000$ A.; this is considered to be due to NaK. It cannot belong to Na_2 which is known not to have any bands in this region, nor can it be attributed to K_2 , for although K_2 has bands in this region, they would not appear without the stronger system of K_2 between $\lambda 4500$ and $\lambda 4200$, and this is absent from the spectrum. The spectrum was finally photographed in the first order of a 21-ft. concave grating. The strong bands lie in the region $\lambda 3870$ – $\lambda 4010$ A., while fainter ones extend to $\lambda 3820$ A. and to $\lambda 4080$ A. The bands are all degraded to the red.

The wavelengths measured are given in table 3. The iron lines were used for comparison. Intensities given are visual estimates on a scale of 10.

Walter and Barratt's measurements include bands down to $\lambda 3540$ A. attributed to NaK; their measurements indicate that there are two strong patches of bands, one between $\lambda 3550$ and $\lambda 3650$ A. and the other between $\lambda 3820$ and $\lambda 3990$ A. Of these, only the latter has been observed in the present investigation and attributed to NaK; in the former region, no bands except those of Na_2 have been observed. It is possible that the strong patch of bands between $\lambda 3550$ and $\lambda 3650$ noted by Walter and Barratt as NaK bands might in fact belong to Na_2 . This suggestion receives support from the fact that all their bands in this region can be fitted in the vibrational scheme of the first ultra-violet system of Na_2 bands and they lie over the Franck-Condon parabola. Table 4 gives the classification of these bands. Weizel and Kulp's arrangement does not include these bands at all. Uchida has included them in his arrangement but the intensity distribution, as given by him, does not appear satisfactory.

§4. VIBRATIONAL ANALYSIS AND DETERMINATION OF MOLECULAR CONSTANTS

The quantum numbers assigned to the ultra-violet bands are given in tables 3 and 5. Nearly all the bands measured can be represented by the equation

$$\nu = 25\,201 + 95.85(u') - 0.94(u'^2) - 123.29(u'') + 0.40(u''^2),$$

where $u = v + \frac{1}{2}$, and the terms in u'' are those of Loomis and Arvin (1934); the differences between the observed values and those calculated from the above equation are given in table 3.

Table 3. NaK bands in absorption: $\lambda 4080$ – $\lambda 3820$.

λ_{lit} (Å.)	Int.	ν', ν''	$\nu_{\text{obs}} - \nu_{\text{calc}}$	λ_{lit} (Å.)	Int.	ν', ν''	$\nu_{\text{obs}} - \nu_{\text{calc}}$
4079.60	1	4.9	0	3907.91	10	{ 4.0 10.4	-3 -1
74.48	1	3.8	3	06.31	4	17.8	-5
62.91	1	0.5	3	03.83	4	12.5	-3
60.61	1	4.8	-1	01.85	4	9.3	-5
55.43	1	3.7	2	00.29	4	14.6	-3
51.49	1	2.6	-2	3898.39	6	16.7	-6
47.21	2	1.5	3	96.33	5	11.4	-1
41.08	2	4.7	1	94.26	8	8.2	1
37.05	2	3.6	2	93.05	4	13.5	-3
32.24	2	2.5	2	89.21	3	{ 10.3 15.6	0 3
28.12	2	1.4	-4	88.62	8	{ 7.1 17.7	-1 -5
23.52	1	0.3	2	85.80	8	12.4	-5
17.20	2	3.5	0	83.85	3	—	—
13.02	2	2.4	-3	82.14	3	{ 6.0 9.2	1 4
08.49	2	1.3	0	81.48	4	14.5	2
04.00	2	0.2	-2	80.13	6	16.6	-3
3997.85	2	3.4	1	79.19	4	11.3	-1
93.62	1	2.3	-2	76.21	8	8.1	-1
88.63	3	1.2	5	75.15	6	13.4	-5
84.33	4	0.1	-1	71.51	4	{ 10.2 15.5	-2 1
78.71	4	3.3	2	70.05	4	{ 7.0 17.6	0 0
74.12	4	2.2	0	68.90	2	19.7	3
69.47	4	1.1	1	67.80	3	12.3	-5
64.93	4	{ 0.0 4.3	0 0	66.17	2	—	—
59.90	6	3.2	0	63.96	4	{ 9.1 14.4	3 0
54.86	6	2.1	0	61.96	3	16.5	-1
51.36	4	5.3	0	59.94	3	11.2	0
45.80	4	4.2	-2	58.26	4	8.0	-3
43.12	2	{ 15.9 10.6	1 8	56.96	4	13.3	-5
40.25	5	{ 3.1 12.7	4 -2	54.00	3	15.4	-6
37.75	2	6.3	2	3852.91	3	10.1	0
35.68	4	2.0	0	51.79	3	17.5	4
34.31	4	—	—	49.93	1	—	—
31.90	3	{ 5.2 8.4	4 -3	48.73	2	—	—
27.00	4	4.1	2	46.73	2	{ 9.0 14.3	-3 -4
24.85	5	7.3	6	43.83	2	16.4	2
21.19	4	{ 3.0 12.6	5 1	42.36	2	11.1	-3
19.82	4	{ 6.2 9.4	-2 -3	40.60	2	—	—
18.47	4	14.7	-5	39.08	1	—	—
16.21	3	16.8	-6	36.04	2	15.3	0
13.59	3	11.5	-5	34.04	1	{ 12.2 17.4	2 5
13.15	4	8.3	1	32.08	1	—	—
12.40	2	—	—	28.55	1	—	—
10.55	4	13.6	0	25.81	1	18.4	-2
08.80	4	—	—				

The heat of dissociation D' can be calculated from Birge and Sponer's extrapolation rule. Its value for the upper state is 2320 cm^{-1} . The value for the ground state, D'' , as given by Loomis and Arvin is 5025 cm^{-1} . With

Table 4. Bands between $\lambda 3550$ and $\lambda 3650$, reported by Walter and Barratt as due to NaK, arranged here as part of the first ultra-violet system of Na_2

λ_{air} (Å.)	Int.	ν', ν''	$\nu_{\text{obs}} - \nu_{\text{calc}}$	λ_{air} (Å.)	Int.	ν', ν''	$\nu_{\text{obs}} - \nu_{\text{calc}}$
3538	1	1,8	-5	3603	2	4,14	15
3545	1	3,10	-5	3607	1	0,11	3
3553	2	0,8	-8	3612	4	2,13	8
3561	2	3,11	10	3622	6	5,16	-4
3569	4	0,9	12	3632	4	2,14	5
3577	4	2,11	0	3641	4	6,18	13
3586	4	5,14	2	3650	2	2,15	-6
3595	2	2,12	2	3654	1	5,18	0

The values of ν_{calc} have been obtained from the equation representing the Na_2 ultra-violet bands between $\lambda 3600$ and $\lambda 3200$ measured recently on a 21-ft. concave grating (unpublished).

Table 5. Deslandres' scheme for the ultra-violet bands of NaK

ν'' ν'	0	1	2	3	4	5	6	7	8	Mean diff.
0	25214	25091	24968	24847		24606				94.3
1		25185	25064	24943	24819					92
2	25400	25278	25156	25033	24912	24793	24675			92.7
3	25495	25372	25246	25127	25006	24886	24764	24652	24536	87
4	25582	25458	25336	25214				24739	24620	89
5	25671	25548	25426	25300						82.8
6	25752	25632	25504	25388						82.5
7	25832	25709	25592	25473						78.8
8	25911	25791	25671	25548	25426					78.8
9		25873	25752	25622	25504					76
10		25947	25822	25704	25582					74.6
11		26018	25900	25778	25658					72.9
12			25975	25852	25727	25609	25495	25372		70
13				25920	25798	25680	25565			70.8
14				25989	25870	25756	25632	25512		69.4
15				26061	25940	25822	25701	25582		67
16					26009	25886	25768			66
17					26075	25955	25832			
Mean diff. (Loomis & Arvin)	122.3 (122.9)	122.1 (122.1)	121 (121.3)	121 (121.4)	118.7 (119.7)	119.4 (118.9)	118.5 (118.1)	117.5 (117.3)		

$\nu_{0,0} = 25215 \text{ cm}^{-1}$, the energy of the dissociated atoms in the excited state, given by $\nu_{\text{atom}} = \nu_{0,0} + \mathbf{D}' - \mathbf{D}''$, is equal to 22510 cm^{-1} . Since $\text{K}4^2\text{S} - 5^2\text{S} = 21025 \text{ cm}^{-1}$ and $\text{K}4^2\text{S} - 3^2\text{D} = 21530 \text{ cm}^{-1}$, and states of Na or other states of K are still farther apart than these, we may reasonably conclude from the above value of ν_{atom} that the dissociated atoms in the excited state are $\text{Na}3^2\text{S}$ and $\text{K}5^2\text{S}$ or 3^2D . It is satisfactory to note that the value of ν_{atom} calculated from the observed data is higher than the value required from considerations of atomic levels. This can be understood, because the value of \mathbf{D}' from the Birge-Sponer extrapolation is generally higher than the true value of heat of dissociation for such molecules (Gaydon 1946). It is, however, difficult to decide whether the K atom after dissociation in the upper level is in 3^2D or 5^2S state. If we assume that the above values of $\nu_{0,0}$ and \mathbf{D}'' are correct, then in order that the K atom should be in 3^2D or 5^2S state, the value of \mathbf{D}' should be 1345 cm^{-1} or 835 cm^{-1} respectively. Of these, in view of the above calculated value of \mathbf{D}' , 2320 cm^{-1} , the former seems more plausible. Hence, it may be concluded that the products of dissociation are $\text{Na}3^2\text{S} + \text{K}3^2\text{D}$ atoms.

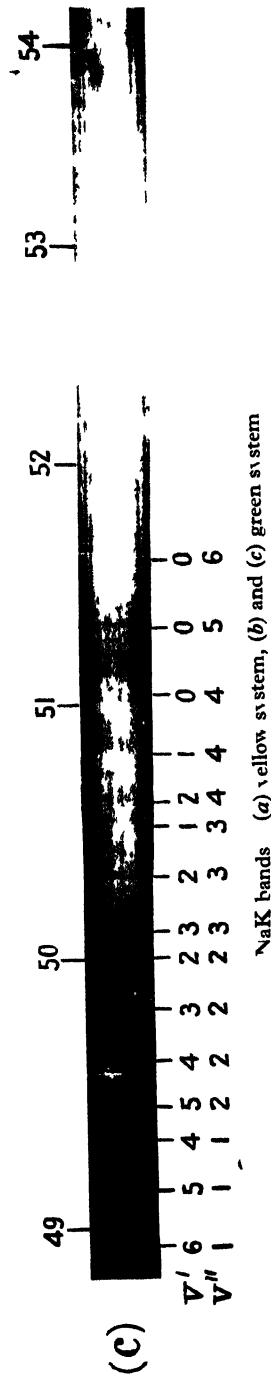
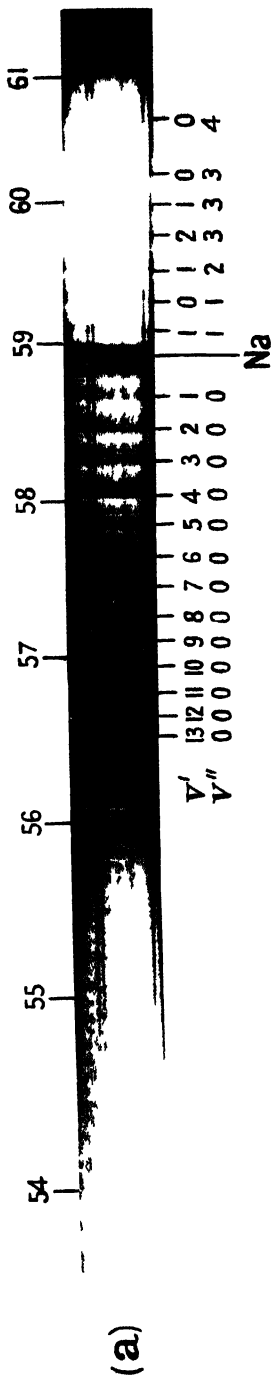
§5. DISCUSSION

The ground state of the NaK molecule will be due to $\text{Na}3^2\text{S}$ and $\text{K}4^2\text{S}$ atoms, and will be a $^1\Sigma^+$ state. Excited states will be obtained if we consider either the sodium or the potassium atom to be raised to higher energy levels. The energy levels of Na and K beginning from the lowest atomic level are in order as follows: $\text{Na}3^2\text{S}$ or $\text{K}4^2\text{S}$, $\text{K}4^2\text{P}$, $\text{Na}3^2\text{P}$, $\text{K}5^2\text{S}$, $\text{K}3^2\text{D}$, $\text{Na}4^2\text{S} \dots$. Ordinarily we expect that in molecules like Na_2 , K_2 or NaK, the lowest molecular state dissociates into atoms in their lowest states, and the next higher molecular state into atoms in the next higher excited states and so on. Hence the molecular states resulting from normal and excited atoms of Na and K will lie roughly in the following order: $^1\Sigma^+$, $^3\Sigma^+$ (from $\text{Na}3^2\text{S} + \text{K}4^2\text{S}$); $^1\Sigma^+$, $^3\Sigma^+$, $^1\Pi$, $^3\Pi$ (from $\text{Na}3^2\text{S} + \text{K}4^2\text{P}$); $^1\Sigma^+$, $^3\Sigma^+$, $^1\Pi$, $^3\Pi$ (from $\text{Na}3^2\text{P} + \text{K}4^2\text{S}$); $^1\Sigma^+$, $^3\Sigma^+$ (from $\text{Na}3^2\text{S} + \text{K}5^2\text{S}$) and $^1\Sigma^+$, $^3\Sigma^+$, $^1\Pi$, $^3\Pi$, $^1\Delta$, $^3\Delta$ (from $\text{Na}3^2\text{S} + \text{K}3^2\text{D}$) In transitions from the ground state, which is $^1\Sigma^+$, we are concerned only with the $^1\Sigma^+$ and $^1\Pi$ states. Comparing the latter with similar states in Na_2 , K_2 or known state of NaK, we conclude that the $^1\Sigma^+$ will be lower than the $^1\Pi$ state. We should, therefore, have the following systems of bands beginning from the long wavelength side:

- Systems 1 and 2: $^1\Sigma^+$, $^1\Pi$ ($\text{Na}3^2\text{S} + \text{K}4^2\text{P}$) $\leftarrow ^1\Sigma^+$
 Systems 3 and 4: $^1\Sigma^+$, $^1\Pi$ ($\text{Na}3^2\text{P} + \text{K}4^2\text{S}$) $\leftarrow ^1\Sigma^+$
 System 5: $^1\Sigma^+$ ($\text{Na}3^2\text{S} + \text{K}5^2\text{S}$) $\leftarrow ^1\Sigma^+$
 Systems 6 and 7: $^1\Sigma^+$, $^1\Pi$ ($\text{Na}3^2\text{S} + \text{K}3^2\text{D}$) $\leftarrow ^1\Sigma^+$

Of these, systems 1, 2 and 4 are the infra-red, the orange and the green systems studied by Loomis and Arvin (1934). System 3 has not been observed and is believed to be overlapped by strong Na_2 bands. It does not seem possible to say whether system 3 would lie at wavelengths longer than those of system 2 or at shorter wavelengths. Loomis and Arvin seem inclined to think that the excited $^1\Sigma^+$ state of system 3 is fairly deep and that the bands of this system would, therefore, lie in the red region where they are obscured by red bands of Na_2 .

The present system of bands can correspond either to system 5 or to system 6. The energy states of $\text{K}5^2\text{S}$ or $\text{K}3^2\text{D}$ are rather too close together to permit



(a) yellow system, (b) and (c) green system

of an unambiguous decision with regard to the dissociation products of the upper state of this system. If we, however, consider the probable error involved in D' we arrive at the conclusion (as shown before) that the dissociated atoms are Na $3^2S + K 3^2D$. This will imply that the bands corresponding to system 5, which will lie at wavelengths longer than those for system 6, are missing. If we consider that like other $^1\Sigma^+$ states, the excited $^1\Sigma^+$ state involved in system 5 is fairly deep so that the bands of this system will lie somewhere in the green region, it is possible that these NaK bands may not be observed at all, having been obscured by strong bands of the green system of Na_2 .

ACKNOWLEDGMENTS

The author wishes to acknowledge his indebtedness to Assistant Professor R. W. B. Pearse, D.Sc., for his kind help and valuable suggestions, and to Patna University, India, for the grant of their Birla Scholarship.

REFERENCES

- GAYDON, A. G., 1946, *Proc. Phys. Soc.*, **58**, 525.
 LOOMIS, F. W., and ARVIN, M. J., 1933, *Phys. Rev.*, **44**, 126 ; 1934, *Ibid* , **46**, 286.
 RITSCHL, R., and VILLARS, D., 1928, *Naturwissenschaften*, **16**, 219.
 SINHA, S. P., 1948, *Proc. Phys. Soc.*, **60**, 436.
 UCHIDA, Y., 1929, *Jap. J. Phys.*, **5**, 145.
 WALTER, J. M., and BARRATT, S., 1928, *Proc. Roy. Soc. A*, **119**, 257.
 WEIZEL, W., and KULP, M., 1930, *Ann. Phys., Lpz.*, **4**, 971.

An Extension of Kapitza's Theory of δ -radiation

By R. M. SILLITTO

The University, Edinburgh

Communicated by N. Feather ; MS. received 20 August 1947

ABSTRACT. The Kapitza theory of δ -radiation is modified by the introduction of a parameter b specifying the radius of the initial temperature distribution. It is shown that a value for b inferred from the columnar ionization theory of Jaffé leads to estimates of the number of δ -electrons, most probable energy of the δ -electrons, and the duration of the emission process, which are in satisfactory agreement with experimental results for platinum. It is also shown that this satisfactory agreement holds only for quite a small range of values of b .

§ 1. INTRODUCTION

KAPITZA'S theory (1923) of the δ -radiation from solid bodies bombarded by α -particles supposes that the energy lost by an α -particle traversing a solid produces a local heating of the body, that the diffusion of this heat through the body can be described in terms of macroscopic constants (e.g. the thermal conductivity and specific heat), and that those parts of the surface which are heated emit electrons according to Richardson's thermionic equation. These thermionic electrons are supposed to constitute the δ -radiation.

It is assumed by Kapitza that the initial heat is localized along a mathematical line coincident with the track of the α -particle. If we denote distance from this line by r , the initial temperature distribution is thus

$$(T)_{t=0} = \infty, \quad r=0;$$

$$(T)_{t=0} = 0, \quad r > 0.$$

The temperature distribution at time t , obtained from the equation of heat conduction (assuming purely radial transfer), is expressed in terms of r , the specific heat c , the thermal conductivity k , the density d , and the energy lost per centimetre by the α -particle Q . In principle the macroscopic constants c , k and d should not be used to describe phenomena localized in a space small compared with the spacing of the atoms in the solid; also it would be preferable to assume that the initial temperature distribution occupies a finite volume. Such a modification of the analysis becomes obviously necessary if, in Kapitza's treatment, we write Richardson's equation in the form $N = A' T^2 e^{-\phi/T}$ instead of $N = AT^{\frac{1}{2}} e^{-\phi/T}$; for otherwise, with this change, the number ν of δ -electrons per α -particle becomes

$$\nu = -(A' Q^2 / 8\pi k c d) \text{Ei}(0) = \infty.$$

The argument of the exponential integral function is proportional to the negative reciprocal of the maximum temperature occurring at $r=0$; for ν to be finite this temperature must be finite, and hence the initial distribution must occupy a finite volume.

§ 2. THE TEMPERATURE DISTRIBUTION

In Jaffé's theory (1913) of the columnar ionization produced in fluids by α -particles, it is assumed that the initial ion density has a cylindrically symmetrical distribution proportional to $\exp(-r^2/b^2)$. This assumption has been justified on theoretical grounds by Lea and Kara-Michailova (1940) who show that b is independent of the energy of the ionizing particle. It is found experimentally that the distribution parameter b is, to a first approximation, inversely proportional to the density of the fluid, and independent of its nature. We shall assume that the initial temperature distribution produced by the passage of an α -particle through a solid* is also of the form $\exp(-r^2/b^2)$, and, further, that an estimate of the magnitude of b can be obtained from the proportionality relation which applies for fluids. It thus appears that for platinum (see below) we may take $b \approx 7 \times 10^{-8}$ cm, or 7 Å., a value which is perhaps large enough to allow the macroscopic constants to have some relevance. (It should be observed, however, that in the application of the columnar ionization theory to fluids it is found that the numerical results obtained are very sensitive to variation of b and the practice is to choose a value of b which leads to results in agreement with experiment. In the discussion which follows it will appear that the result of the calculation of the number of δ -electrons is similarly very sensitive to variations of b .)

* There must obviously be a certain, though extremely short, time lag between the passage of the α -particle and the establishment of a physically recognizable temperature distribution; it is perhaps in the recognition of this point that the method of the present paper differs significantly from that of Kapitza.

For the initial temperature distribution we now write

$$(T)_{t=0} = T_0 + (Q/\pi cdb^2) \exp(-r^2/b^2) \quad \dots\dots(1)$$

where T_0 is the temperature of the target before the entry of the α -particle. It follows that, at time t ,

$$T = T_0 + \frac{Q}{\pi cd(4kt/cd + b^2)} \cdot \exp\left(-\frac{r^2}{4kt/cd + b^2}\right). \quad \dots\dots(2)$$

§ 3. THE NUMBER OF δ -ELECTRONS

If the α -particle enters the solid normal to the surface, and we assume that (2) gives the temperature distribution on the surface, the total δ -emission is obtained by combining (2) with

$$\nu = \int_0^\infty \int_0^\infty \pi A' T^2 e^{-\phi/T} dr^2 dt \quad \dots\dots(3)$$

and evaluating the integral. Following Kapitza we change this to an integral over surface elements $d\omega$ in the (r^2, t) plane for which T is constant. The analysis is considerably simplified if we drop the constant T_0 of equations (2) and (3), and consider the range of temperature variation as being from 0 to T_{\max} (say) rather than from T_0 to $T_0 + T_{\max}$. If T_0 represents room temperature, say 300°K . and T_{\max} is of the order of 10^4°K . (see below), no significant error is introduced by this approximation. We then find that

$$d\omega = \left\{ \frac{Q^2}{8\pi^2 kcdT^3} - \frac{cdb^4}{8kT} \right\}$$

and hence

$$d\nu = \frac{A'}{8k} \left\{ \frac{Q^2}{\pi cd} \frac{e^{-\phi/T}}{T} - \pi cdb^4 T e^{-\phi/T} \right\} dT. \quad \dots\dots(4)$$

When $t=0$, $r=0$, we have $T = T_{\max} = T_m = Q/\pi cdb^2$, so that

$$\nu = \frac{A'Qb^2}{8k} \int_0^{T_m} \left\{ \frac{T_m}{T} \cdot e^{-\phi/T} - \frac{T}{T_m} \cdot e^{-\phi/T} \right\} dT. \quad \dots\dots(5)$$

If $\theta = -\phi/T$, $\theta_m = -\phi/T_m$, (5) becomes

$$\nu = \frac{A'\phi Qb^2}{8k} \int_{-\infty}^{\theta_m} \left(\frac{1}{\theta_m} - \frac{\theta_m}{\theta^2} \right) \frac{e^\theta}{\theta} d\theta$$

$$\text{giving} \quad \nu = \frac{A'\phi Qb^2}{8k} \left\{ \left(\frac{1}{\theta_m} - \frac{\theta_m}{2} \right) \text{Ei}(\theta_m) + \left(1 + \frac{1}{\theta_m} \right) \frac{e^{\theta_m}}{2} \right\}. \quad \dots\dots(6)$$

For the α -rays from polonium falling at the beginning of their range on a platinum target, $Q \simeq 2 \times 10^{-10}$ cal/cm. For Pt, du Bridge gives $A' = 1.02 \times 10^{23}$ electrons/sec. $\text{cm}^2 \text{deg}^2$, and $\phi = 7.28 \times 10^4$ °K. Taking $c = 0.05$ cal/gm., corresponding to a temperature in the neighbourhood of 2×10^3 °K., we find $T_m = 1.21 \times 10^4$ °K. From (6), therefore, we obtain $\nu = 11$ electrons/ α -particle.

Experimental values for ν lie between about 0.5 and 15; a value between 5 and 10 seems most probable.

§ 4. THE ENERGY OF THE δ -ELECTRONS

As indicated by Kapitza, (4) may be regarded as specifying the "distribution in temperature" of the emitted electrons. The equation may be rewritten as

$$d\nu = \frac{A'Qb^2}{8k} \left\{ \frac{T_m}{T} e^{-\phi/T} - \frac{T}{T_m} e^{-\phi/T} \right\} dT \quad \dots\dots(4')$$

of electric dipole and quadrupole transitions. For the magnetic dipole transitions only the total number of pairs regardless of angle is given as a function of E' .

I have carried out a similar calculation for the angular distribution in the case of magnetic dipole transitions. The number of pairs for which θ lies within an interval $d\theta$ and in which E' lies between $\epsilon' mc^2$ and $(\epsilon' + d\epsilon') mc^2$ is found to be:

$$P(\epsilon', \theta) d\epsilon' \sin \theta d\theta = 2\pi\gamma(\epsilon' - 1 - \mathbf{p}' \cdot \mathbf{p}'') \left\{ \frac{(\epsilon' \epsilon'' + 2) - 3\gamma^{-2}(\epsilon' \epsilon'' + 1)^2}{- \mathbf{p}' \cdot \mathbf{p}'' [1 - 4\gamma^{-2}(\epsilon' \epsilon'' + 1)]} - \gamma^{-2}(\mathbf{p}' \cdot \mathbf{p}'')^2} \right\} d\epsilon' \sin \theta d\theta,$$

where

$$\alpha = \frac{e^-}{hc}, \quad |\mathbf{p}'| = (\epsilon'^2 - 1)^{\frac{1}{2}}, \quad |\mathbf{p}''| = (\epsilon''^2 - 1)^{\frac{1}{2}}$$

$$\mathbf{p}' \cdot \mathbf{p}'' = |\mathbf{p}'| |\mathbf{p}''| \cos \theta, \quad \gamma = h\nu/mc^2,$$

$h\nu$ being the disintegration energy and $\epsilon'' = \gamma - \epsilon'$ is the energy of the positron in units of mc^2 . E' and E'' include the rest energy.

As it is difficult to determine experimentally the total number of gamma quanta emitted by an assembly of excited nuclei, it is not easy to measure the absolute probability of internal pair production per emitted gamma quantum (Jaeger and Hulme 1935). Another type of experimental procedure would be to measure the number of pairs as a function of the angle θ between the two components of the pair. A comparison with the calculated results, discussed in this paper, will then show the multipolarity of the gamma ray. In order to facilitate this comparison we give graphs for the three cases, magnetic and electric dipole and electric quadrupole transitions.

Figure 1 shows, for two given total energies, the number of pairs produced per disintegration, in equal intervals of $\cos \theta$, as a function of $\cos \theta$. Equal intervals of $\cos \theta$ correspond to constant solid angle and are convenient for counting experiments in which the solid angle subtended by each counter will usually be independent of the relative position of the two counters.

Figures 2 and 3 show the number of pairs produced, per disintegration, against $\cos \theta$, for given electron-positron energies. We have shown only two representative cases, one in which the electron carries half the energy, the other in which one carries most. It is easy to see from these two cases the qualitative behaviour for other energy distributions.

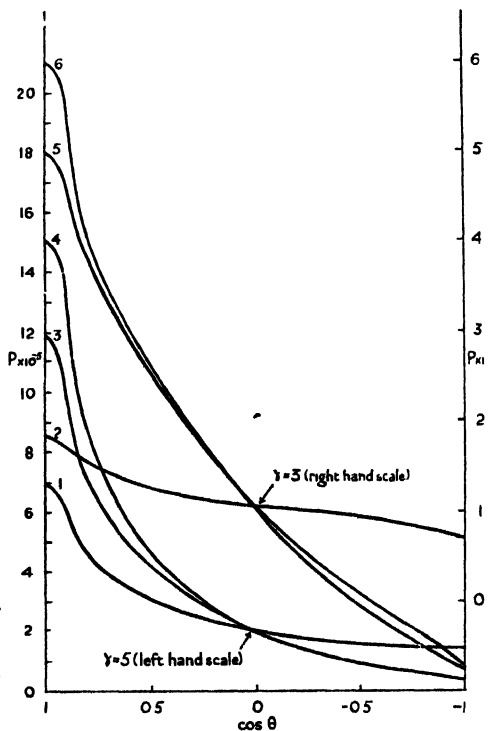


Figure 1. Overall angular distribution of pairs.

- 1, 2, Electric dipole.
- 4, 6, Magnetic dipole.
- 3, 5, Electric quadrupole.

The absolute values are given for the electric dipole case and the other two normalized to coincide with it at $\cos \theta = 0$.

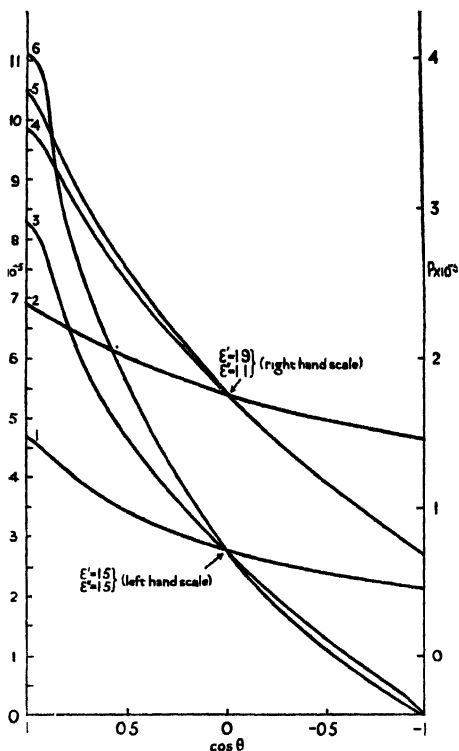


Figure 2. The probability for a given energy distribution with $\gamma = 3$.

- 1, 2, Electric dipole.
- 5, 6, Magnetic dipole.
- 3, 4, Electric quadrupole.

The absolute values are given for the electric dipole case and the other two normalized to coincide with it at $\cos \theta = 0$.

Lastly we show, in figure 4, the average value of $\cos \theta$ plotted against various energy distributions among the components of the pair and for two values of the incident photon energy.

The following conclusions may be drawn. We have obtained results symmetrical in the electron and positron by neglecting the electrostatic field of the nucleus. The magnitude of the effect is of the order $0.5 - 1.5 \cdot 10^{-4}$ pairs per nuclear disintegration. The angular distributions are qualitatively similar. Let us call K_D and K_Q respectively the ratios of the number of pairs per disintegration at $\cos \theta = 1$ to those at $\cos \theta = -1$ in

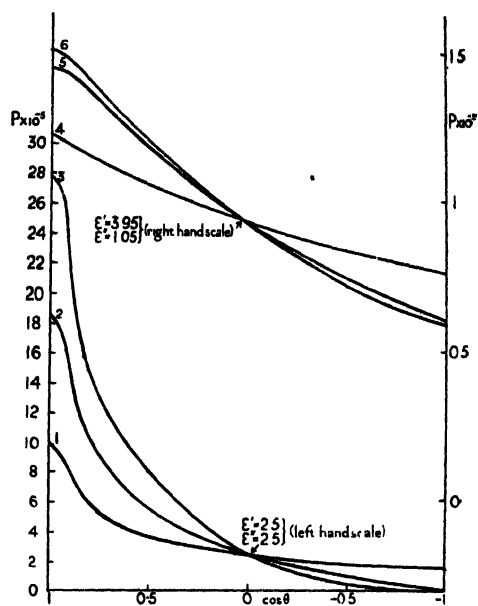


Figure 3. The probability for a given energy distribution with $\gamma = 5$.

- 1, 4, Electric dipole.
- 3, 5, Magnetic dipole.
- 2, 6, Electric quadrupole.

The absolute values are given for the electric dipole case and the other two normalized to coincide with it at $\cos \theta = 0$.

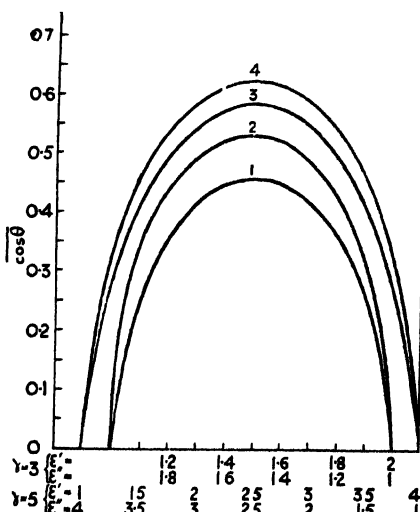


Figure 4. The average of $\cos \theta$ for $\gamma = 3$ and 5.
1, 3, Magnetic dipole or electric quadrupole.
2, 4, Electric dipole.

the case of electric dipole and quadrupole (or magnetic dipole) transitions respectively and let $K = K_Q/K_D$. (We see from the curves that electric quadrupole and magnetic dipole transitions are hardly distinguishable by this method.)

In figure 1, for $\gamma = 5$, $K \simeq 8$ and for $\gamma = 3$, $K \simeq 15$. In figures 2 and 3, for very asymmetric energy distributions between the components of the pair, we find $K \simeq 2$. When the electron and positron share the available energy equally, we find that $K \simeq 5 \cdot 10^3$ for γ equal to both 3 and 5.

In figure 4 we find that the average angle between the pair, for $\gamma = 3$, is the same for both electric quadrupole and magnetic dipole radiation. In both these cases the average angle, for each distribution of energy among the components of the pair, is about 20% larger than in the case of electric dipole transitions. The same applies for $\gamma = 5$ when all the average angles are somewhat smaller.

The method may be used, together with selection rules such as parity and angular momentum, to interpret nuclear spectra by assigning angular momentum quantum numbers to the various levels.

ACKNOWLEDGMENT

The author wishes to thank Professor R. E. Peierls for his guidance of this work and the Department of Scientific and Industrial Research for a maintenance grant.

REFERENCES

- JAEGER, J. C., and HULME, H. R., 1935, *Proc. Roy. Soc. A*, **148**, 708.
 OPPENHEIMER, J. R., and NEDELSKI, L., 1933, *Phys. Rev.*, **44**, 948 ; corrected in 1934, *Ibid.*, **45**, 283.
 ROSE, M. E., and UHLENBECK, G. E., 1933, *Phys. Rev.*, **48**, 211.

The Correction for Self-Weakening in β -Ray Measurements *

By E. BRODA¹, W. E. GRUMMITT², J. GUÉRON³, L. KOWARSKI³
 AND G. WILKINSON⁴

¹ Now at Department of Natural Philosophy, The University, Edinburgh.

² Now at National Research Council, Chalk River, Ontario. -

³ Now at Commissariat à l'Energie Atomique, Paris.

⁴ Now at Chemistry Department, University of California.

MS. received 18 November 1947

ABSTRACT. Experimental details are given for the determination of self-weakening of β -rays in thick samples. The self-weakening correction factor in the case of twelve radio-elements is found over a considerable region to depend logarithmically on the mass-thickness of the sample. The self-weakening half-thicknesses $\eta_{\frac{1}{2}}$ is as a rule 2.2 times larger than the external weakening half-thickness $\epsilon_{\frac{1}{2}}$. The logarithmic dependence of the self-weakening correction factor on mass-thickness and the value of the ratio $\eta_{\frac{1}{2}}/\epsilon_{\frac{1}{2}}$ are derived in a simplified theoretical treatment.

* The work described in this paper was done for the Department of Scientific and Industrial Research, and the National Research Council, Canada, during 1942 and 1943. It was reported at that time, but could not be released for publication due to security regulations.

§ 1. INTRODUCTION

THE measurement of the specific activity, i.e. of the number of disintegrations per unit time and per unit mass, of a given β -ray emitting sample is frequently involved in the exploration of neutron densities with detectors, and in the determination of cross-sections for nuclear reactions leading to the formation of β -emitters. Such measurements are usually carried out by placing the sample close to a β -ray sensitive device (Geiger counter or electroscope), which, however, fails to respond to such β -particles as may be absorbed or scattered away during their passage through the sample itself. This "self-weakening" has to be taken into account as a correction when interpreting the response of the measuring device.

The necessity for this correction can sometimes be circumvented. If the sample is thin enough, self-weakening may be neglected altogether, but the recourse to thicker samples is often necessary in order to get enough intensity. Many kinds of comparative measurements, in arbitrary units, can be carried out by using samples of exactly identical thickness, so that the self-weakening corrections cancel out. The preparation of samples of exactly prescribed thickness is, however, inconvenient, especially if the material is short-lived and needs chemical treatment before counting. Moreover, this device is of no avail if different β -emitters are to be compared.

It might be thought that the "true" specific activity (i.e. that of a sample of vanishing thickness) could be deduced from the observed specific activity (of a thick sample) by applying rules on β -ray absorption, using values of absorption coefficients predetermined in suitably chosen external screens. But absorption and scattering of β -rays are so complex, and the sample-and-counter geometry is so little amenable to mathematical treatment, that a purely empirical approach seems desirable. Broda, Guéron and Kowarski (1942) calibrated the observed specific activity n for various mass-thicknesses of several β -emitters (^{238}U , ^{239}U , ^{56}Mn and ^{128}I) in samples of equal area and in their particular counting geometry. A linear relationship between the mass-thickness and the logarithm of the specific activity was found to hold in a wide range, including in particular the domain of small thicknesses. With the help of such calibration curves the true specific activity, i.e. the specific activity as measured with very thin samples, can be deduced from the value n measured with thick samples. The same authors listed values of the "mass-thickness for self-weakening to one-half" ($\eta_{\frac{1}{2}}$) i.e. the mass thickness at which the β -ray intensity per unit weight of a given sample is reduced to one-half. Grummitt, Guéron and Wilkinson (1943) investigated the same elements in a different counting geometry. They confirmed the general conclusions drawn by Broda, Guéron and Kowarski, but their values of $\eta_{\frac{1}{2}}$ were slightly different, perhaps because of the changes in the geometry. Grummitt, Guéron and Wilkinson also extended the method to ^{233}Th , ^{139}Ba , ^{140}Ba , ^{140}La , and ^{131}I . Results of other authors (Hendricks, Bryner, Thomas and Ivie 1943, Yankwich, Rollefson and Norris 1946, and Cook, private communication) on ^{35}S , ^{14}C and ^{232}Th will be shown to be amenable to a similar treatment. While the method, therefore, is widely applicable, the numerical values quoted in this paper are valid only under the given experimental conditions.

§ 2. EXPERIMENTAL PROCEDURES

(i) *Chemical*

(a) *Mn 56 and I 128.* These activities were obtained by exposure to slow neutrons of manganese oxide (empirical formula $\text{MnO}_{1.47}$) and lead iodide. The activity of these compounds was measured without chemical treatment after irradiation.

(b) *U 239.* Uranium as sodium uranyl acetate was irradiated with slow neutrons. In order to reduce the natural β -ray activity, and to remove fission products, chemical treatment was necessary. Excess sodium acetate was added with stirring to a cold solution containing 10% uranium by weight and 25% acetic acid by volume. The crystalline precipitate of sodium uranyl acetate was washed on the filter with saturated sodium acetate solution, a small amount of water, and finally with acetone. It was then dissolved in the minimum amount of nitric acid, the concentrations of uranium and acetic acid were adjusted and the uranium reprecipitated. Two sodium uranyl acetate precipitations before irradiation reduced the UX activity almost to zero; after irradiation, a further two precipitations reduced the fission product activity and newly grown UX activity to less than 1% of the initial U 239 count. The samples were ready for counting 10-15 minutes after the end of irradiation.

(c) *Th 233.* Thorium also had to be freed from its decay products. The procedure was to add to a solution of thorium nitrate before irradiation barium and lanthanum carriers (for radium and actinium) and to precipitate their carbonates with excess ammonium carbonate. The filtrate, containing the soluble thorium carbonate complex, was acidified with nitric acid, and after addition of more barium salt barium was precipitated again with sulphuric acid and removed by filtration. Now lanthanum, bismuth, lead and barium ions were added as hold-back carriers, and the thorium was precipitated at 60° C. with excess 40% H_2O_2 as thorium peroxynitrate. This was redissolved in the minimum quantity of boiling concentrated HNO_3 with the addition of a little HCl. The peroxynitrate precipitation, again in presence of a hold-back carrier, was repeated, and the precipitate ignited to ThO_2 . The residual β -activity of non-irradiated ThO_2 was about 40 counts per minute per gramme (counter efficiency about 25%).

(d) *Ba 140 and La 140.* These activities were separated from the mixture of fission products formed in the slow neutron irradiation of uranium oxide or uranyl nitrate. The details of the chemical procedure (Grummitt, Guéron, Wilkinson and Yaffe, 1947) follow in general the fuming nitric acid method of Hahn and Strassmann (1942).

(ii) *Preparation of samples and counting*

At first cylindrical aluminium counters of wall thickness 40 mg/cm² were used. The counters were filled to 4 cm. Hg with argon and to 2 cm. Hg with alcohol vapour. The finely powdered activated materials were spread evenly on rectangular aluminium trays of area 2.1 cm². A few drops of an acetone solution of collodion (≈ 0.5 mg/cm² collodion) were added as binder, the sources dried by gentle heating and covered with a single layer of "sellotape". The sources were disposed around the central portion of the counter tube. This.

method allows efficient utilization of weak sources due to the large area exposed. It is, however, an inconvenient and inaccurate arrangement for work with external absorbers.

In the later measurements (Grummitt, Guéron and Wilkinson 1943) a "bell jar" type of β -ray counter with a mica window of area 7.1 cm^2 and mass thickness 5 mg/cm^2 was used. It was filled to 9 cm. Hg with argon, and to 0.5 cm. Hg with alcohol. Circular aluminium trays of area 7.1 cm^2 were supported 1 cm. below the window by means of a sample holder placed in a slotted frame. Samples were prepared as previously, but were not covered with "sellotape".

In no case were the samples counted without mechanical or chemical mixing after irradiation. This is important as the effect of unequal activation at different depths, due to the absorption of the effective neutrons, particularly resonance neutrons, has to be excluded.

(iii) Measurements

The measured activities of the samples were corrected for variations in the counter efficiency with the help of uranium standard. In the cases of U 239 and Th 233 an additional correction of the measured β -ray counts was necessary due to the growth of the natural β -ray activity in the samples after chemical purification. Correction was made with U 239 by counting the UX activity some time after the decay of the 23.5 minutes period with extrapolation back to the time of counting of U 239. With Th 233 a non-irradiated reference sample was used; since the fission products of Th were too weak to interfere, chemical treatment after irradiation was not necessary.

The results were expressed in terms of n , the apparent specific activity of the material, i.e. as the number of counts per minute per gramme of active material at an arbitrary reference time. The logarithm of n was then plotted against the mass thickness M of the sample. In figure 1 typical plots for U 239 are given. As the neutron flux varied between the sets of experiments, the results of each set were normalized at a point in the middle of the mass-thickness range studied. Extrapolation of this composite curve to mass thickness zero gives a value n_0 , the true specific activity of the sample. In figure 2 the

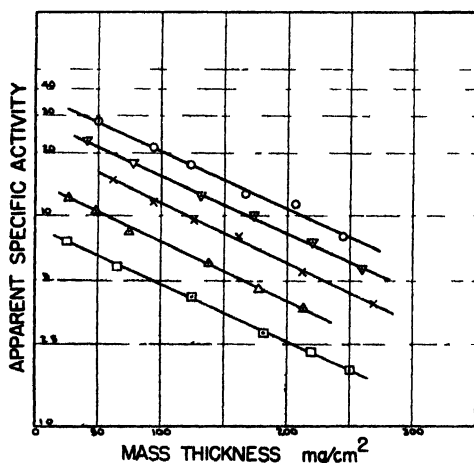


Figure 1. Self-weakening of U239 β -particles in sodium uranyl acetate

self-weakening calibration curves, i.e. the plots of $\log(n_0/n)$ against M , are given for the bodies investigated. The self-weakening correction for any thickness is read directly from these curves. For all the radiations measured $\log n_0/n$ depended linearly on M . In table 1 the values of the self-weakening half-thicknesses η_1 , as read from the calibration curves, are collected. Further, the table gives the weakening half-thicknesses ϵ_1 as determined by measurement of the β -ray absorption in aluminium foils placed midway between sample and

Table 1

Radioactive nucleus	Active material	$\eta_{\frac{1}{2}}$ (mg/cm ²)	$\epsilon_{\frac{1}{2}}$ (mg/cm ²)	$\eta_{\frac{1}{2}}/\epsilon_{\frac{1}{2}}$
UX ₂	Sodium uranyl acetate	243 ^b 228 ^a	135 ^b	1.81
U 239 (23.5 min.)	Sodium uranyl acetate	113 ^b 92 ^a	54 ^b	2.1
Mn 56 (2.54 h.)	MnO _{1.47}	458 ^b 512 ^a	171 ^b	2.74
I 128 (25 min.)	PbI ₂	266 ^b 297 ^a	128 ^b	2.08
Th 233 (23 min.)	ThO ₂	134 ^b	56 ^b	2.4
La 140 (40 h.)	LaF ₃	156 ^b	76 ^b	2.06
Ba 140 (12.7 d.)	BaSO ₄	91 ^b	41 ^b	2.23
S 35 (87.1 d.)	BaSO ₄	6.9 ^c	3.14 ^c	2.2
C 14 (4700 y.)	BaCO ₃	6.0 ^d	3 ^d	2.0
I 131 (8.0 d.)	CuI	56 ^b 60 ^e	22 ^b 25 ^e	2.54 2.4
Ba 139 (86 min.)	BaSO ₄ Ba(NO ₃) ₂	323 ^b 295 ^e	147 ^b	2.2
MsTh 2 (6.13 h.)	BaBeF ₄	145 ^f		

^a Broda, Guéron and Kowarski.^b Grummitt, Guéron and Wilkinson.^c Hendricks, Bryner, Thomas and Ivie 1943.^d Yankwich, Rollefson and Norris 1946.^e Broda 1943, 1944.^f Cock, private communication.

counter. It is to be noted that the external absorption for UX₂, Mn 56 and I 128 in the present geometrical arrangement does not follow an exponential relationship over even a limited portion of the range.

Figure 2 and table 1 also give self-weakening data calculated from experiments by other authors to conform with the $\log n_0/n$ against M presentation as used above. It is noteworthy that even with β -radiations as soft as those of S 35 and C 14 the logarithmic relation holds over a considerable range of mass thicknesses.

In table 1, column 5, the ratio of the self-weakening half-thickness and the weakening half-thickness (with external screens) is listed. With a few exceptions, notably UX₂ and Mn 56, the values of this ratio fall close to the average $\eta_{\frac{1}{2}}/\epsilon_{\frac{1}{2}} = 2.2$.

§ 3. DISCUSSION

The experimental results show that the relationship between the self-weakening correction factor and

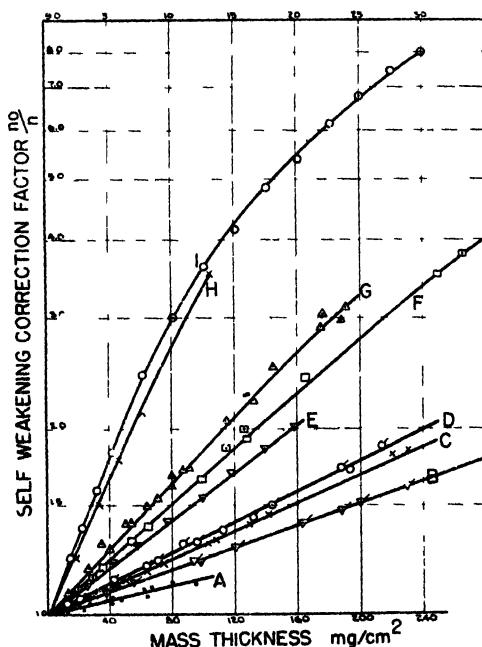


Figure 2. The self-weakening correction factor n_0/n vs mass thickness for the following β -radiations:

- A. Mn56 in manganese oxide (MnO_{1.47}).
- B. Ba139 in barium sulphate.
- C. I128 in lead iodide.
- D. UX₂ in sodium uranyl acetate.
- E. La140 in lanthanum fluoride.
- F. Th233 in thorium dioxide.
- G. U239 in sodium uranyl acetate.
- H. S35 in barium sulphate (upper scale).
- I. C14 in barium carbonate (upper scale).

the mass-thickness of the active sample can be considered as logarithmic over a considerable mass range.

The relation can be deduced by a simplified theoretical treatment with the following assumptions :

(a) The measurements are made in a parallel beam geometry.

(b) The weakening of β -rays as measured by external screens obeys an exponential law over the initial portion of the curve, and the weakening is independent of the nature of the screen, i.e. $n = n_0 e^{-\mu x}$ where x is the mass thickness of the absorber, μ the mass absorption coefficient, and n_0 and n the numbers of incident and transmitted particles. The external half-thickness $\epsilon_{\frac{1}{2}}$ is $0.693/\mu$.

(c) The back-scattering of β -rays by the sample and by the backing material can be considered as identical. This assumption is justifiable if backing material of atomic number similar to that of the active material is used.

Then the contribution to the total β -activity of a layer of mass thickness dm at a depth m in a thick sample of active material of specific activity n_0 is

$$dA = n_0 e^{-\mu m} dm.$$

The total activity of the sample of mass thickness M is, therefore,

$$A = \int_0^M n_0 e^{-\mu m} dm = (n_0/\mu)(1 - e^{-\mu M}).$$

The apparent specific activity n of the sample is

$$n = A/M = (n_0/\mu M)(1 - e^{-\mu M})$$

and the ratio of the apparent and the true specific activity is

$$\frac{n}{n_0} = \frac{(1 - e^{-\mu M})}{\mu M} = 1 - \frac{\mu M}{2!} + \frac{\mu^2 M^2}{3!} - \frac{\mu^3 M^3}{4!} + \dots \quad (1)$$

In figure 3 the expression $\log \mu M/(1 - e^{-\mu M})$ is plotted as a function of M . Thus it appears that on the above assumptions the self-weakening correction curve n_0/n against M is approximately logarithmic over the initial region. The value of μM , which reduces the activity function (1) to $\frac{1}{2}$, is 1.53, i.e. $\mu \eta_{\frac{1}{2}} = 1.53$ and $\eta_{\frac{1}{2}} = 2.21 \epsilon_{\frac{1}{2}}$.

Serious divergence from the linear relationship occurs at values of $\mu M > \sim 2$, i.e. at values $M > \sim 3 \epsilon_{\frac{1}{2}}$. The experimental correction curve for C 14 in barium carbonate shows divergence from linearity at mass thicknesses greater than 8 mg/cm^2 , i.e. approximately 2.7 times the external absorption half-thickness for the β -rays of C 14. However, in the case of S 35 no deviation is evident up to $M = 4.5 \epsilon_{\frac{1}{2}}$.

Feather (1943, 1944) has submitted the whole problem of self-weakening of β -rays to a careful theoretical treatment, and has given

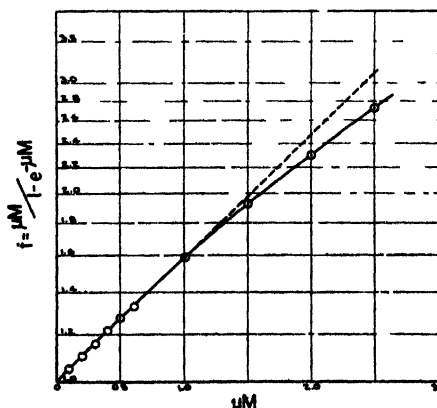


Figure 3. The theoretical self-weakening relation as given by a simplified treatment.

explicitly the conditions under which a logarithmic relation, as found in our work, may be expected to hold.

It is to be expected from the well-known complexities of β -ray scattering and absorption that the numerical values will vary with the geometrical conditions of measurement. However, the relation $\eta_1 = 2.2\epsilon_1$ is approximately obeyed in the majority of the experiments listed in table 1. Such discrepancies as exist are not accounted for by the differences between the average atomic number of the active material and the external absorber (Al), nor is there so far any obvious correlation with the energies of the radiation. Data on radiations from more radioelements are necessary before drawing any conclusion.

REFERENCES

- BRODA, E., 1943, *British Atomic Energy Report*, BR. 365, declassified as BBDA. 45 ; 1944, *Ibid.*, BR. 506, declassified as BBDA. 57.
 BRODA, E., GUÉRON, J., and KOWARSKI, L., 1942, *British Atomic Energy Report*, BR. 44.
 FEATHER, N., 1943, *British Atomic Energy Report*, BR. 315, declassified as BBDA. 114 ; 1944, *Ibid.*, BR. 504
 GRUMMITT, W. E., GUÉRON, J., and WILKINSON, G. 1944, *Canadian Atomic Energy Report*, MC. 46.
 GRUMMITT, W. E., GUÉRON, J., WILKINSON, G., and YAFFE, L., 1947, *Canadian J. Res.*, **25B**, 364 (*Canadian Atomic Energy Report*, MC. 58).
 HAHN, O., and STRASSMANN, F., 1942, *Naturwissenschaften*, **30**, 325.
 HENDRICKS, R. H., BRYNER, L. C., THOMAS, M. O., and IVIF, J. O., 1943, *J. Phys. Chem.*, **47**, 469.
 YANKWICH, P. E., ROLLEFSON, G. K., and NORRIS, T. H., 1946, *J. Chem. Phys.*, **14**, 131.

The β -Ray Spectrum of ThC''D.

By D. G. E. MARTIN, H. O. W. RICHARDSON * AND YUN-KUEI HSÜ

George Holt Physics Laboratory, University of Liverpool

* Now at the Department of Natural Philosophy, University of Edinburgh

MS. received 10 July 1947

ABSTRACT. The energies of the β rays of thorium C''D have been measured using a magnetic spectrograph with semi-circular focusing. Corrections have been made for contamination of the source by ThC and ThB deposited by the processes of β -recoil and aggregate recoil. The absolute intensities of the stronger lines have been measured and the shape of the continuous spectrum has been compared with theory using three recently proposed energy level schemes for the product nucleus Pb²⁰⁸.

These level schemes fail to explain the observed shape, partly because they omit the P γ -ray (859 kev) which can now be assigned to ThC''D. Evidence for the β -excitation of energy levels at about 4.5 mev is discussed.

§ 1. INTRODUCTION

SEVERAL level schemes have been proposed for the nucleus ThD. Among the more recent are those of Oppenheimer (1936), Arnoult (1939) and Itoh and Watase (1941). It is of interest to compare the β -ray spectra predicted by these schemes with the β -ray spectrum of ThC''D found by experiment.

The short half life of 3.1 min. of ThC'' indicates that the β -disintegration belongs to the class of allowed transitions. The spectrum is the sum of a number of partial spectra having different end points and intensities and each exciting a different level of the product nucleus ThD (Pb^{208}). All the stronger partial spectra should be allowed transitions and should have the shape predicted by the Fermi theory.

The shape of the experimental spectrum found in the present work deviates from those predicted by the schemes hitherto proposed for the energy levels and γ -ray intensities of the product nucleus Pb^{208} . Much of this deviation may be due to the omission from the schemes of the 859 kev γ ray whose assignment of the $\text{ThC}''\text{D}$ disintegration has been confirmed by our measurements. A surplus of low energy β rays suggests that one or more levels of Pb^{208} at about 4.5 mev receive β -excitation.

Caution is required in the interpretation of the observations because of the presence of ThB and C nuclei carried over in the preparation of the ThC'' sources by α -recoil. Corrections for this contamination have been subtracted. The thin mounting used for the source should not have distorted the spectrum above 100 kev.

Alichanian and Zavelskij (1937) found relatively more β rays at both high and low energies than were found in the present work.

§ 2. THE EXPERIMENTAL METHOD

The sources of ThC'' were deposited on the thin aluminium leaf A (figure 1) by α -recoil in vacuo from a source of thorium ($\text{B} + \text{C} + \text{C}''$). The latter, which may be called the activating source, was withdrawn when counting. The short half life of ThC'' necessitated the use of many recoil sources, whose intensities were correlated by measuring their γ -ray activities with an independent γ -ray counter in a fixed position.

The deposit of ThC'' was found to be contaminated with thorium B and C, which also emit β rays. This contamination was due to the processes of aggregate recoil and β -recoil. In the former, an atom of ThC'' recoiling strongly from the emission of the α ray in the ThC'' decay, carries with it a random

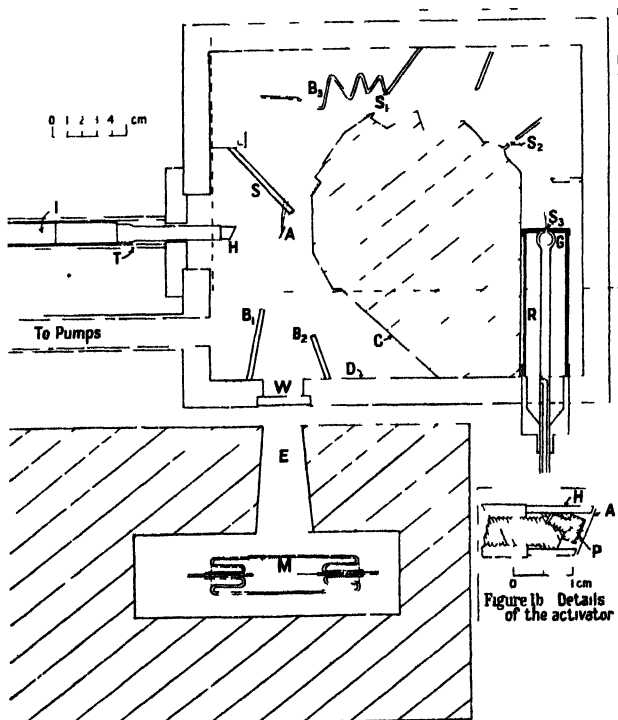


Figure 1. The magnetic spectrograph.

sample of neighbouring atoms from the surface of the activating source. In the second process a small fraction of ThC atoms leaves the surface of the activating source owing to the weak recoil momentum received from the emission of the decay electron and neutrino from ThB.

The aggregate recoil effect was the more serious, and its magnitude was followed in each day's run by repeated measurements of the peak of the very strong F β -ray line (1386 $H\rho$) of thorium B.C. The magnitude of this effect showed wide variations on different days.

§ 3. THE MAGNETIC SPECTROGRAPH

During the preparation of the 'ThC'' sources recoil particles may be deposited on other surfaces in the spectrograph in addition to the source mounting A. By closing the shutter S (figure 1) during the time when the activating source was near to A, recoil particles were prevented from settling near the main defining slit S_1 or from passing through it. If S was left open during the process of activation the background counting rate of the β -counter showed erratic variations. These were probably due to charged β -recoil atoms of low momentum deviated through the slits by the magnetic field. These atoms of ThC contribute to the count at zero magnetic field when they disintegrate near the counter.

The baffle system shown in figure 1 was designed to exclude from the counter β rays emitted from these deposits on the surfaces of the spectrograph and also scattered β rays. β rays coming from atoms on surfaces other than A and lying on the source side of the main slit S_1 were excluded from the counter slit S_3 by the second defining slit S_2 . The surfaces C and D did not, however, lie in the zone of exclusion defined by the slits S_1 , S_2 , and S_3 . β rays from the surface D were excluded by adding baffles B_1 and B_2 of the type used by Neary (1940). β rays from the surface C and from the edges and left sides of B_1 and B_2 were not excluded, but scattering from them was reduced by inclining them so that only their edges were exposed to the source. This should also keep them free from α -recoil particles, which, owing to their great momentum, travel in straight lines in the magnetic field. The baffle B_3 was designed to keep secondary aggregate recoil particles (coming from A when the shutter S was open) away from a region where the thickness of lead screening the β -counter was small.

The aluminium leaf A was 9 mm. wide and 3.5 cm. long and had a weight of 0.13 mg/cm². It was inclined at 22° to the central ray of the focused beam of radius 9 cm. The depth and the greatest width of the beam at S_1 were both 1 cm. A high vacuum was maintained by using a 2-stage oil diffusion pump.

The plane pole-faces of the magnet were rectangular in form of sides 19 × 28 in. and were 5.9 cm. apart. Their edges are shown by the broken line in figure 1. The magnet showed the phenomenon of magnetic viscosity and, in order to get a reproducible relation between magnetizing current and field strength, readings were always taken on the ascending branch of the same hysteresis cycle which had a constant maximum field of 1400 oersteds. The field was calibrated by using a flip-coil and ballistic galvanometer. To get an absolute scale of $H\rho$ we assigned to the field settings at the centroids of the β -ray lines the values of $H\rho$ found for the lines by Ellis (1932). This procedure, discussed by Lawson and Tyler (1940), is correct if the curvature of the continuous spectrum is small. (The shape of the F-line of thorium B.C is shown in figure 2 (a). The width of the line at half its height is about 1.35% of H . The arithmetic mean or centroid

of the line shape lies about 0.2% of $H\rho$ to the left of the middle of the two points half-way up opposite sides of the line.)

§ 4. THE β - AND γ -RAY COUNTERS

The β and γ Geiger-Müller tubes (G and M in figure 1) were filled with mixtures of argon and alcohol vapour. The cathode of the β -counter was of brass 2 cm. long and 1 cm. diameter. The anode was a tungsten wire 0.1 mm. in diameter. The reservoir R ensured constancy of the gas mixture of 6 cm. of argon with 2 cm. of alcohol. The slit S_3 was 2.4 mm. wide and 1 cm. high and was covered with mica.

The use of the γ rays to correlate the sources requires that the ratio of the efficiencies of the two counters must remain the same over an appreciable time. A source of error is introduced by the finite slopes of the plateaux of the characteristic curves of the two counters. The β -counter had a plateau about 60 volts long and a slope of about 0.15% per volt at 930 volts. The γ counter had a slope of 0.4% per volt at 1500 volts. The voltages were stabilized by small neon tubes and were kept as constant as possible with respect to the lower end of the plateaux.

On each day the weak ends of the spectrum were counted first, the middle parts being left until the source was weaker. In this way a day's run roughly covered the whole spectrum so that day-to-day changes in relative efficiency should introduce no systematic distortion. This method of sandwiching the measurements should make random the error due to small changes in voltage.

The dependence of the efficiency of the β -counter on the energy of the β rays was found by measuring the spectrum of thorium ($B + C + C''$) in equilibrium, at different pressures in the counter. The efficiency at high energies (935 kev) did not begin to fall until the pressure of argon was below 6 cm. with 2 cm. of alcohol. It was concluded that at least 98% of the β rays of $\text{ThC}''\text{D}$ were counted at all energies.

The loss of counts due to the finite resolving time of the counting systems was considered to be negligible. Each counter fed into a Kipp-relay equalizing circuit followed by a pentode scale-of-eight circuit due to O. R. Frisch. The maximum counting rates were about 1200 per minute.

The absorption of slow β rays in the β -counter window was estimated by measuring the spectrum of $\text{Th}(B + C + C'')$ and comparing it with the results of Flammersfeld (1939), who used a much thinner window. The surface density of the mica was about 2 mg/cm², β rays begin to penetrate at 600 $H\rho$, while the intensities of the B and B β lines at 653 and 678 $H\rho$ were respectively 19% and 30% of Flammersfeld's values. At 1000 $H\rho$ the difference between the two spectra was only about 10%, a loss which is much less than that found, for example, by Martin and Townsend (1939). This is probably because the part of the counter slit outside the mica window was narrower than the part inside it, so that rays could suffer appreciable sideways deflection and still be counted. The conditions resembled those in Schonland's experiments (1925) and by fitting his law of absorption for aluminium to the present measurements we estimate the fraction of the incident beam lost in the mica window to be 10%, 2.6%, and 0.6% at $H\rho$ 1085, 1542 and 2186, respectively. These corrections have not been included in figure 4. They would increase the downward convexity of that end of the spectrum.

§ 5. PREPARATION AND CORRELATION OF THE SOURCES

The activating source was obtained by overnight exposure to thoron of a platinum disc P, 4.5 mm. in diameter, which was then screwed into the "activator" shown in figure 1*b*. A high α -recoil efficiency was obtained by polishing P with chromic oxide and exposing it to the thoron in air at 6 cm. pressure in order to reduce tarnishing.

The hood H was pushed on to the activator to prevent the sideways escape of recoil particles from P, which, if they fell near the glass window W of the spectrograph, would give a disproportionately large γ -count. The gap between H and A was about 0.5 mm.

In order to reduce the effect due to γ -ray sources not on the leaf A, the γ -counter was entirely surrounded by a lead shield 7.3 cm. thick, whose only opening was the collimating channel E, 2.7 cm. wide and 0.8 cm. deep at its outer end.

The usual working cycle was 14 minutes; of this 6½ minutes were used for counting and the rest for activation. In order to obtain a reliable measure of the source strength it was arranged that the γ -ray count was more than 2½ times the β -ray count, even at the peak of the continuous spectrum.

During counting the activator was withdrawn about 1.3 metres down the curved tube T by means of a sliding magnet acting on an iron rod I. When withdrawn an additional 24 cm. of lead separated the activator from the counters, making the effect of its γ -radiation negligible.

Tests with a fluxmeter showed that the motion of the activator magnet altered the magnetization of the main magnet and that owing to magnetic viscosity the latter took about 1 minute to reach a steady state. It was therefore necessary to go round the standard hysteresis cycle during each activation.

§ 6. THE CORRECTION FOR AGGREGATE RECOIL

Aggregate recoil is the only recoil process which may deposit thorium B on the source mounting A. It may thus be studied separately by measuring the strong F line which contains more than 0.25 electrons per disintegration and stands out about 30 times higher than the maximum of the thorium B continuous spectrum at the resolving power of our apparatus. The growth in each day's run was followed by making about four measurements of β_F , the counting rate at the peak of this line.

The counting rate β_A due to aggregate recoil at a given setting of the magnetic field H can be calculated if β_F and β_A/β_F are known. Using small greek letters for rates and capitals for total counts, we have

$$\begin{aligned} B &= B'' + B_0 + B_B + B_A \\ &= B'' + [\beta_0 + \beta_B + \beta_F(\beta_A/\beta_F)]\Delta t, \end{aligned} \quad \dots\dots(1 a)$$

where B is the total β -particle count in a counting period of length Δt (usually 6½ minutes); B'' is the count due to ThC''D with half-life 3.1 min.; β_0 is the natural background rate in the β -counter, and β_B is the rate due to β -recoil. β_B decays with the 1-hour period of Th(C+C'') and thus may be considered constant during Δt .

The contribution Γ_A of the aggregate recoil to the γ -count was also proportional to β_F . We have

$$\begin{aligned} \Gamma &= \Gamma'' + \Gamma_0 + \Gamma_B + \Gamma_A \\ &= \Gamma'' + (\gamma_0 + \gamma_B + 0.03\beta_F)\Delta t, \end{aligned} \quad \dots\dots(1 b)$$

where 0.03 is the experimental value found for γ_A/β_F .

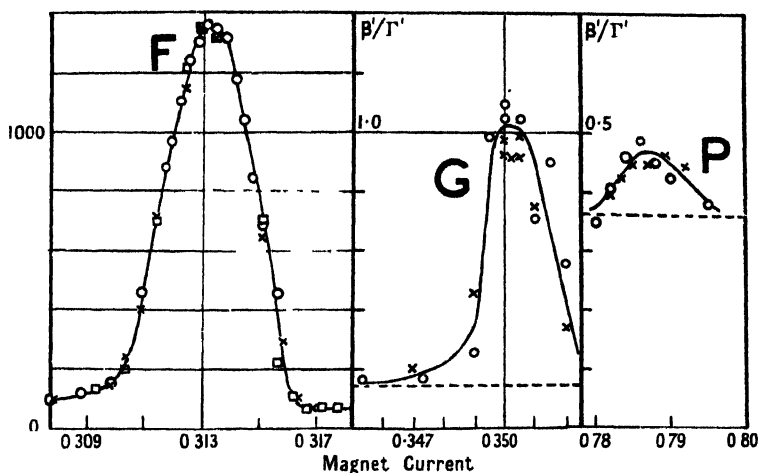


Figure 2. β -ray lines. The F line of ThB.C by aggregate recoil and the G and P lines of $\text{ThC}''\text{D}$ by α -recoil. The broken lines indicate the continuous spectrum.

From these equations B''/Γ'' , the standardized β -ray intensity of $\text{ThC}''\text{D}$, was found for each value of H . The measurement of β_F was made easier by the flatness of the peak of the F line, as seen in figure 2. The flat part had a width of about 0.5% of $H\rho$ and the field was increased through this zone in steps of about 0.2%. β_F was taken as the highest reading lying between two lower ones, all three containing more than 2000 β -counts. This choice of the highest of three readings subject to fluctuation might give rise to a systematic error. Analysis showed that the small overestimation of β_F from this cause was, by a happy chance, nearly cancelled by the mean fall in the highest reading due to its distance from the exact top of the line. This mean fall increases with the spacing of the field settings and the curvature of the peak.

The aggregate recoil efficiency fell during the course of each day, in agreement with the findings of Lawson (1919). It showed big variations from source to source. In order to find B''/Γ'' by equation (1), it is necessary to know β_F during each counting period Δt . β_F was estimated by the method of graphical interpolation shown in curve (1) of figure 3, where B_n and A_{n+1} are values of β_F found from the n th measurement of the F-line. In a group of activations running for example from A_1 to B_1 , a little more than half the time was

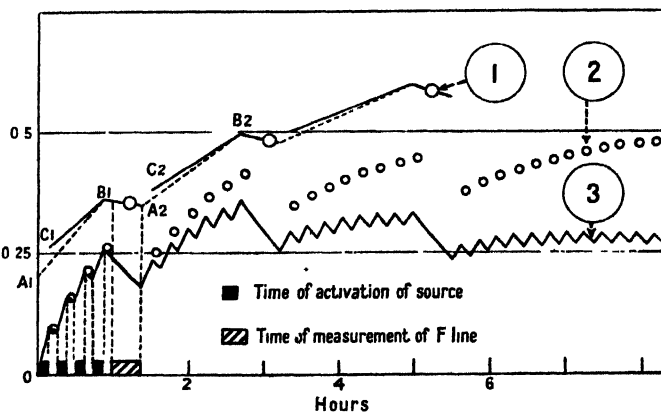


Figure 3. Growth of contamination by aggregate and β -recoil. (1) Counting rate $\times 10^{-3}$ at the peak of the F line of ThB.C . (2) n/ne , the fraction of the equilibrium strength of the β -recoil contamination. (3) $(n/ne)e^{-\lambda t}$, proportional to its absolute strength n .

spent in activation so that the mean rate of growth of β_F during activation was just less than twice the slope of A_1B_1 . Since, however, this rate was always falling, a good compromise can be made by taking the rate during the first activation A_1C_1 to be twice the slope of A_1B_1 and then assuming a linear growth from C_1 to B_1 . The values of β_F are then read off along C_1B_1 , C_2B_2 etc.

The values (β_A/β_F) in equation (1) were found in two steps. First, one value of this ratio was found by counting at the maximum of the aggregate recoil spectrum on days when it was strong, applying a correction for β -recoil. The second step was to fit this known value of the ratio at $H\rho=4541$ to the known spectrum of thorium ($B+C+C''$) in equilibrium on a thin aluminium leaf, from which other values of β_A/β_F could then be read. The stepwise comparison was necessary because the relative height of the F-line above the continuous spectrum was not necessarily the same in the two spectra, the height and shape of a β -ray line being sensitive to changes in the position and distribution of the source on A. No such changes occurred during the measurement of $ThC''D$ because activation by recoil does not disturb A and seems to give a constant surface distribution.

The errors in the values of ΣB_A given in table 1 arise (i) from the use of only 2000 counts to find β_F , (ii) from the graphical interpolation, (iii) from the value of (β_A/β_F) at $H\rho=4541$ and (iv) from the spectrum of $Th(B+C+C'')$ in which about 30 000 counts per point were used. It is thought that the errors are covered by assigning a standard deviation of 3% to ΣB_A .

Table 1. The continuous spectrum, the corrections for aggregate and β -recoil and values of the Fermi function $F(Z, \eta)$

$H\rho$	ΣB	ΣB_0	ΣB_A	ΣB_B	$\Sigma B''/\Sigma I''$	% $\sigma(N)$	$F(Z, \eta)$ for $Z=82$
0	1,542	1017	0	0	0.0078		
1085.5	22,455	2756	6254	306	0.0808	2.1	3.208 ₃
1541.6	32,261	2835	8548	547	0.1489	1.7	4.906 ₁
2185.8	21,228	1597	2735	468	0.2390	1.2	7.678 ₁
2770.1	17,906	1115	2049	433	0.3082	1.3	10.667 ₈
3306.7	20,839	1169	2847	635	0.3496	1.3	13.77 ₁
3843.5	19,674	1155	2443	595	0.3714	1.3	17.20 ₁
4541.2	18,259	1176	2163	603	0.3313	1.3	22.11 ₈
5077.9	20,582	1409	2840	813	0.2663	1.3	26.23 ₈
5594.7	25,767	2096	4579	1267	0.1787	1.5	30.45 ₇
6128.9	23,834	2702	4230	1433	0.1104	1.8	35.08 ₂
6662.4	16,497	2847	2992	1151	0.0513	2.5	39.95 ₄
6928.6	7,894	1776	1433	519	0.0319	3.7	42.47 ₈
7194.4	2,578	1015	462	228	0.0132	16.3	45.05 ₈

$$\eta = H\rho/1704.2$$

ΣB = Total number of β -counts.

ΣB_0 = Natural background.

ΣB_A = Counts due to aggregate recoil.

ΣB_B = Counts due to β -ray recoil.

$\Sigma B''/\Sigma I''$ = Standardized β -ray intensity.

% $\sigma(N)$ = Percentage standard deviation of N (cf. eqn. (5)).

§ 7. THE CORRECTION FOR β -RECOIL

The counts assigned to this effect are shown as ΣB_B in table 1.

The counting rates are so small that in contrast to aggregate recoil no direct method is available for finding the growth of the contamination at intervals during

the day. For simplicity we assume that the ratio of the α - and β -recoil efficiencies is constant for deposition on A from the activator. The value of β_B in equation (1 a) can then be estimated for each counting period Δt .

The first step is to assign to each Δt a measure of the relative strength of the β -recoil source at that time. This measure is given by n/n_e , the ratio of the number n of β -recoil atoms on A to the equilibrium number n_e which would have been reached if the interval had been preceded by a very long period of activation.

Let us assume that a constant fraction β of the ThB atoms decaying on the activator deposits recoil atoms of ThC on A. Then at time t during an activation

$$n = n_i e^{-\lambda_2(t-t_i)} + \beta \frac{\lambda_1 n_0 e^{-\lambda_1 t}}{\lambda_2 - \lambda_1} [1 - e^{(\lambda_1 - \lambda_2)(t-t_i)}], \quad \dots\dots (2)$$

where n_i is the initial number of recoil atoms n on A at the start of the activation at $t = t_i$, n_0 is the initial number of ThB atoms on the activator at $t = 0$, and λ_1 and λ_2 are the decay constants of ThB and ThC. Between activations n decays with decay constant λ_2 . The equilibrium value n_e is given by

$$n_e = n_0 \beta \lambda_1 e^{-\lambda_2 t / (\lambda_2 - \lambda_1)}. \quad \dots\dots (3)$$

Using (2) and (3) average values of n/n_e were estimated for each counting period Δt . An example of the serrated curve of n against t is shown in figure 3 for a typical day's run.

The second step is to find the ratio of the α - and β -recoil efficiencies. Defining the β -recoil efficiency β as in equation (2) and the α -recoil efficiency α as the fraction of ThC.C'' disintegrations which deposits ThC'' on A, let us consider the case of an activation so long that both γ'' and γ_B have reached transient equilibrium and n_i/n_e is unity. Then every N atoms of ThB decaying on the activator will deposit $N\beta$ atoms of ThC on A by β -recoil and $0.34 N\alpha$ atoms of ThC'' by α -recoil (because only 34% of ThC atoms decay to ThC'').

If we neglect the weak γ rays of ThC.C' and ThC.C'' the γ -activity on A will be due to ThC''.D alone and will be the sum of the counting rates γ'' and γ_B due to the two modes of arrival by recoil. The ratio of the numbers of ThC'' atoms arriving by each mode (neglecting the 10% difference between secular and transient equilibrium) is

$$\frac{0.34 N\alpha}{N\beta \times 0.34} = \frac{\alpha}{\beta} = \left(\frac{\gamma''}{\gamma_B} \right)_{\text{equilibrium}}. \quad \dots\dots (4)$$

The ratio of γ'' to γ_B at equilibrium was found by measuring the decay of the activity of strong recoil sources made by long activations. The decay of the β -activity at $H_p = 4541$, and of the γ rays, was followed for several hours and was analysed into three half-periods of 3.1 min., 60.5 min. and 10.6 hours due respectively to α , β and aggregate recoil. Extrapolating back gave γ_B and γ'' , but owing to the weakness of the effect large statistical errors in γ_B were present. Seven measurements of $(\gamma''/\gamma_B)_{\text{equil}}$ on different days gave a mean of 51 with a mean residual of 15%. Five earlier measurements of the decay of the γ -activity, each at the end of a day's run, gave the value 64 with 6% mean residual. As the value of β is known to be sensitive to changes in the emitting surface layer it is possible that the difference is significant and that it arose from a change which was made in the method of making the source. The earlier value was therefore used for measurements taken before the change was made.

The value of γ_B was calculated for each counting period Δt , obtaining γ'' in (4) from the observed γ -count Γ , and using the assigned value of n/n_0 . The value of the β -recoil correction β_B in equation (1 a) was found by multiplying γ_B by the value (β_B/γ_B) appropriate to each value of $H\rho$. This ratio was obtained by direct measurements of the β - and γ -counting rates of 1-hour half-life at $H\rho = 4541$. Its value at other fields was estimated from the known shape of the β -spectrum of $\text{Th}(C + C'')$ deduced from that of $\text{Th}(B + C + C'')$ by subtracting from the four points of lowest energy the contribution due to thorium B. This contribution was estimated from the curves of Flammersfeld (1939), making a small correction for loss of $\text{Th}C''$ by α -recoil from the aluminium.

In view of possible variations of β -recoil efficiency, a standard deviation of 10% is assigned to $\Sigma\beta_B$ in table 1 which makes it the major source of error at the four highest values of $H\rho$.

§ 8. THE β -RAY LINES

Flammersfeld (1939) has found intensities of the β -ray lines in the spectrum of $\text{Th}(B + C + C'')$ which are higher than those previously accepted. The increase arises in two ways. First, the absolute intensity of the F-line (1386 $H\rho$), used ordinarily as a standard, was found to be 0.284 electrons per disintegration as against the former value of 0.250 (Gurney 1926). Second, the intensity of the lines of $\text{Th}C''D$ was increased by about 30%, by covering the source with a thin film to prevent the escape of $\text{Th}C''$ atoms by α -recoil. This loss by recoil may be expected to occur from all uncovered sources which are moderately clean.

It is interesting to check these intensities by a method which is independent both of the F-line strength and of loss by α -recoil. A second point of interest is to verify the assignment of some of the many lines of doubtful origin. A $\text{Th}C''D$ line will be about three times as strong in our case as it is in the $\text{Th}(B + C + C'')$ spectrum, whereas other lines will have reduced intensities set by the amount of the aggregate and β -recoil contamination.

Approximate values for the intensities of the lines are given in table 2. The absolute intensity of each line in electrons per disintegration is equal to the ratio

Table 2. Line intensities in electrons per disintegration

β -ray line	Ec	Fb	G	Jb	Jb3	L	M	N	P	X
$H\rho$ (Ellis)	1251	1469	1594	1929	1980	2603	2887	3164	3924	10 000
Abs. Int. in $\text{Th}C''D \cdot 10^2$	0.12	0.50	2.50	0.35	0.08	1.77	1.49	0.38	0.26	0.18
Int in $\text{Th}(B + C + C'')$	—	0.356	2.73	0.416	—	2.26	1.75	0.356	0.255	0.234
(Flammersfeld)				(Jb + Jb1)						
Gamma ray	Ec	Fb	G	G	G	L	L + M	M	P	X

of the area due to the line to that due to the disintegration electrons in the momentum distribution curve. This curve is obtained by plotting N against $H\rho$ where

$$N = \frac{1}{H\rho} \left(\frac{B''}{\Gamma''} - \left[\frac{B''}{\Gamma''} \right]_0 \right) \quad \dots\dots (5)$$

$[B''/\Gamma'']_0$ is the standardized β -ray intensity when $H=0$ and is due to the counts caused by the γ rays of thorium C'' in the β -counter. The area beneath each line is equal to the product of the height by the mean width. The latter was found to be 0.0137 $H\rho$ by measuring the area of the F-line shown in figure 2.

Flammersfeld's values are given for comparison in the table, taking the thorium C branching ratio to be 33.7% (Kovarik and Adams 1938). The intensities in row 2 are subject to error because it was necessary when measuring a ThC''D line to reset the magnetic field after each activation. It was found impossible to reproduce H to better than about 1 in 300 after going round the standard magnetization cycle. It was thus difficult to build up a large number of counts on the flat peak of a line by repeated activations at the same H . A low β -count due to just missing the peak (about $H\rho/200$ wide) is indistinguishable from a statistical fluctuation. This uncertainty did not arise in the measurement of the F-line of thorium B because there was then no magnetic disturbance due to motion of the activator.

In view of this limitation of the method the ratios of the intensities of the G, L, and M lines were found independently from the spectrum of Th(B+C+C') in transient equilibrium. The lines now stand out less above the continuous spectrum, but it is possible to get about 20 000 counts at each peak.

The relative intensities of the lines are compared with the results of Flammersfeld (1939) and Oppenheimer (1936) in table 3.

Table 3. β -ray line intensity ratios

L : G	M : G	X : G	Source
0.826	0.641	0.0857	Flammersfeld (1939)
0.429	0.353	0.106	Oppenheimer (1936)
0.709	0.597	0.0711	ThC''D data in table 2
0.692	0.619	—	Th(B+C+C')

Special weight must be attached to the L : G ratio of 0.69 found using Th(B+C+C') because it is the mean of three independent measurements. It seems likely that the correct ratio lies between the values of Flammersfeld and Oppenheimer.

The uncertainty in setting the field can be seen from the typical plot of the G and P lines in figure 2. Each point is the value of B'/I' found from a single activation. B' is the β -ray count not corrected for β -recoil nor for the counts due to the line itself in the aggregate recoil spectrum. This correction, when made, improves the agreement between data taken on different days. The purely statistical standard deviation in B'/I' due to the finite numbers counted is about 1.8% for the G line in which about 10 000 β -counts and 12 000 γ -counts were recorded near the peak. In the case of the P line this error is about 10%, and it is higher for the weaker lines. It is likely that the error due to just missing the peak was larger than the statistical error for the stronger lines. It would be expected to lead to an underestimate of the intensity because it would effectively flatten out the lines. Thus the fact that our intensities are rather lower than those of Flammersfeld may be due to a systematic error. They confirm his results to the extent that they are appreciably higher than the values of Ellis (1932) and Oppenheimer (1936).

The reduction in apparent height of a β -ray line due to error in setting the field H may be estimated roughly if we represent the line by a gaussian error curve

$$\beta = \beta_1 e^{-h_1^2 x^2} + \beta_0,$$

where β_0 is the counting rate due to the continuous background, β_1 = the peak height of the line at $H = H_1$, and $x = (H_1 - H)/H_1$. Let there now be added an

independent gaussian error in H with a precision index h_2 . The broadened line must have the same area so we get

$$\beta' = \beta_1(h/h_1)e^{-h^2/h_1^2} + \beta_0$$

where $h^2 = h_1^{-2} + h_2^{-2}$. If we take $h_1 = 1.23 \times 10^3$ and $h_2 = 3.54 \times 10^2$, equivalent to percentage standard deviations of 0.572 and 0.2 respectively, we get a decrease in the height β_1 of the line of 6%.

Small drifts in the field strength due to magnetic viscosity also lead to an underestimate of the height of the lines.

A search was made for the lines recorded by Meitner and Philipp (1933) as having moderate intensity. Their line with $H\rho = 541$ is below the limit for penetration of the counter window, while that with $H\rho = 1379$, presumably the E_{α} line, is too close to be separated from the F-line of thorium B in our case. These authors did not survey the region containing the P-line which was attributed to $^{232}\text{ThC}'''\text{D}$ by Sze (1933). The P γ -ray must carry an appreciable fraction of the total γ -ray energy.

A weak line was found near $H\rho = 1990$. This was probably the Jb3 line of Ellis (1932) which he attributed to the internal conversion of the Ga γ -ray of thorium CC'' in the L_1 level. However, its energy is also close to that expected for the conversion of the strong G γ -line of $^{232}\text{ThC}'''\text{D}$ in the M_1 level. The intensity we find for this line when using $^{232}\text{ThC}'''\text{D}$ sources supports its assignment to the G γ -ray, but it may well be a case of a line with a double origin such as the M line. Surugue (1937) found two close lines at $H\rho$ 1778 and 1981 which he attributed to the conversion of the Ga γ -ray in the L_1 level and the G γ -ray in the M_1 level respectively.

§ 9. THE CONTINUOUS SPECTRUM

The momentum distribution for an allowed β -transition (Fermi 1934) may be written

$$\sqrt{(P(\eta)/F(Z, \eta))} = A^{1/2}(\epsilon_0 - \epsilon), \quad \dots\dots(6)$$

where $P(\eta)d\eta$ is the probability of emission of a β -ray with momentum lying between η and $\eta + d\eta$ and ϵ_0 is the maximum value of ϵ , the energy of the β -ray including its rest-energy, expressed as a multiple of mc^2 . We have used $mc^2 = 510.89$ kev and $\eta = H\rho/1704.2$.

For a given value of the atomic number Z , A is proportional to the square of the matrix element of the heavy particle transition in the nucleus

$$F(Z, \eta) = \eta^{2S} \cdot e^{\gamma\epsilon} \cdot |\Gamma(S + i\gamma\epsilon, \eta)|^2 \quad \dots\dots(7)$$

where $\gamma = Z/137$, and $S = \sqrt{(1 - \gamma^2)}$.

$F(Z, \eta)$ has been computed with $Z = 82$ for 13 values of $H\rho$ and is given in table 1. It is believed that the values of F are correct to four significant figures.

$P(\eta)$ is proportional to N , defined in equation (5). The graph of $\sqrt{(N/F)}$ against ϵ is given in figure 4 for values of ϵ chosen to avoid the β -ray lines. The errors to be expected in the values of N are expressed as percentage standard deviations in table 1. These errors were found by assigning zero error to Γ_0 and Γ_B and standard deviations of 1%, 3%, 10% and 5%, to B_0 , B_A , B_B and Γ_A respectively. The uncertainty rises rapidly near the end-point at $\epsilon = \epsilon_0$.

The graph deviates from the straight line predicted by (6) for a single β -transition, which suggests that the spectrum is the sum of several partial β -spectra ending on different energy levels of the product nucleus Pb^{208} . All the stronger

partial spectra should be allowed and should have the definite shape predicted by the Fermi theory. At low energies the shape predicted may no longer be definite because it depends on the coefficients in the combination of terms which may be present in the expression for the interaction in β -decay. Evidence on this has been given by Lewis and Bohm (1946) for Cu^{64} , but no clear indication of the effect appears in the present work.

It can be seen from figure 4 that the end-point at $\epsilon = \epsilon_0$ is approached almost linearly. If we adopt the view discussed in the next section that the experimental end-point corresponds to the β -excitation of an energy level 3.20 mev above the ground state of Pb^{208} , we cannot expect another strongly excited level until a

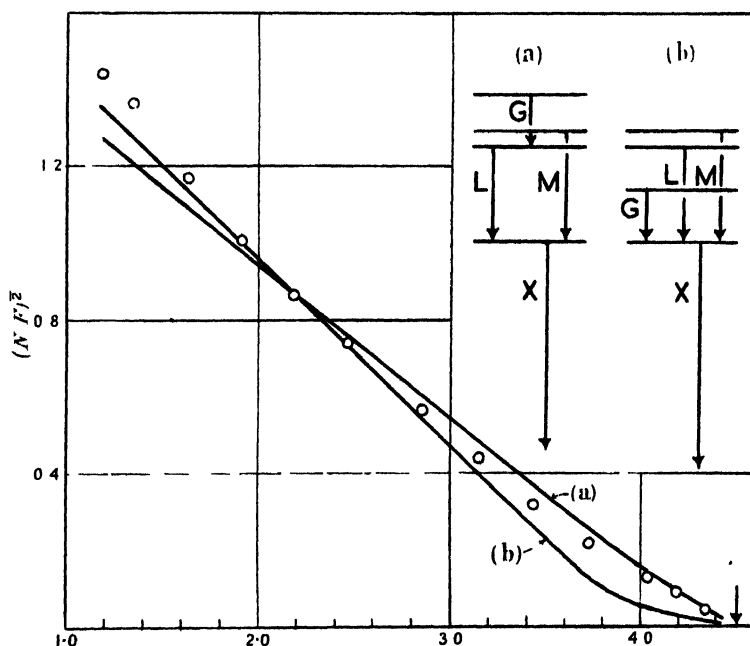


Figure 4. Comparison of the experimental spectrum with the spectra calculated from (a) the level scheme of Itoh and Watase and (b) from that of Oppenheimer.

space equal to the energy of the G γ -ray of 277 kev has been bridged. Thus the last 277 kev of the spectrum should be linear, but below this there should be a small rise in slope. The consequences of such a level scheme are discussed elsewhere, but we may use it provisionally to estimate ϵ_0 by the method of least squares by fitting the four points of highest energy ϵ to a straight line.

The deviations of the four points were weighted in the normal equations inversely as the standard deviations of N given in table 1. The value of $(N/F)^{1/2}$ at the point with $\epsilon = 3.73 \text{ mc}^2$ was reduced by 4% because it should receive a small contribution from the second partial spectrum. We get for the maximum energy $\epsilon_0 = (4.507 \pm 0.014) \text{ mc}^2$, or $(1.792 \pm 0.007) \text{ mev}$. This agrees with the value of 1.795 mev found by Henderson (1934) using a method which avoided the deposition of aggregates on the source, presumably by carrying out the activation in air.

The probable error was calculated by means of the standard formula (Margenau and Murphy, p. 502). It depends only on the deviations of the four points and does not take account of a possible error in our scale of $H\rho$ and may thus be an underestimate.

§ 10. COMPARISON OF ALTERNATIVE ENERGY LEVEL SCHEMES

If the downward convexity of the spectrum in figure 4 is due to its being the sum of a series of m partial β -spectra all having the shape of an allowed transition, it should obey the equation

$$\chi^2(N, F) = \chi^2 \left(\sum_{n=1}^m A_n (\epsilon_n - \epsilon)^2 \right) \quad \dots\dots(8)$$

where $A_n = 0$ for $\epsilon > \epsilon_n$.

A_n is proportional to the square of the heavy particle matrix element for the partial β -transition to the n th energy level of the product nucleus, with end-point at $\epsilon - \epsilon_n$.

A_n can be connected with p_n , the β -transition probability to the n th level, which is usually called the β -excitation.

We have

$$p_n = \int_0^{\epsilon_n} p(\epsilon) d\epsilon \quad \dots\dots(9)$$

where $p(\epsilon)d\epsilon$ is the probability of emission of a β -ray in the n th partial spectrum with energy between ϵ and $\epsilon + d\epsilon$. To evaluate the integral we choose as an approximation for $F(Z, \eta)$ in (6).

$$F(Z, \eta) = 3.85\eta(1 + 0.418\eta). \quad \dots\dots(10)$$

This gives, on integrating (9)

$$p_n = KA_n \left[\frac{1}{12} \epsilon_n^4 - \frac{1}{2} \epsilon_n^2 + \frac{2}{3} \epsilon_n - \frac{1}{4} + (0.418(\epsilon_n^2 - 1))^{\frac{1}{2}} \left(\frac{1}{30} \epsilon_n^4 - \frac{1}{20} \epsilon_n^2 - \frac{12}{15} \right) + 0.418 \cdot \frac{1}{4} \epsilon_n \ln (\epsilon_n + (\epsilon_n^2 - 1)^{\frac{1}{2}}) \right]. \quad \dots\dots(11)$$

K is a constant which is adjusted to make $\sum_1^m p_n = 1$ in a given level scheme.

Equation (10) differs from the formula of Fermi (1934) in that 0.418 replaces 0.355. We believe that the value 0.418 covers the energy range which makes the main contribution to p_n in our case.

Among the more recent schemes proposed for the energies and β -excitations p_n of the levels of thorium D are those of Oppenheimer (1936), Arnould (1939) and Itoh and Watase (1941). The second scheme differs from the first partly in that it includes a number of weak γ -rays some of which are of doubtful origin. It would give much the same shape of spectrum. Using (8) and (11) we have calculated the shapes of composite spectrum predicted by the first and third schemes, taking the largest end-point, ϵ_1 , to be $4.50 mc^2$. The spectra shown in figure 4 are fitted to the experimental curve at $\epsilon = 2.177$ by adjusting K . The values of the constants are given in table 4.

Table 4. The energy level schemes and the resulting complex β -spectra

Author	Oppenheimer			Itoh and Watase		
Energy above ground level (MeV)	2.895	3.13	3.20	3.13	3.20	3.41
γ -ray transitions to ground level	G + X	L + X	M + X	L + X	M + X	G + L + X
β -ray end point ϵ_n	4.50	4.04	3.90	4.50	4.36	3.96
β -excitation p_n	0.08	0.20	0.73	0.45	0.45	0.10
KA_n	0.169	0.722	3.199	0.949	1.16	0.399

The discrepancies between the spectra seem to be beyond the limits of experimental error and are too great to be removed by any changes in the values of p_{β} , which would be compatible with the intensities of the γ -ray transitions. Upper and lower limits to these intensities can be deduced from the theory of internal conversion using the observed intensities of the β -ray lines.

A difficulty in both these schemes is that they do not put the end-point of the β -spectrum on the energy level at 3.20 mev. It was pointed out by Henderson (1934) that in order that the total energy emission may be equal round both sides of the ThC branching it is necessary to place the β -ray end-point at 1.792 mev on or near the 3.20 mev level of ThD . Otherwise one must postulate different rest-masses for the neutrinos emitted in ThC''D and in ThCC' . In order to give a correct balance of the branching energies Arnould (1939) included in his level scheme a partial spectrum with an excitation p_{β} of 10% extending to an end-point at 2.10 mev. No β -ray evidence has been found for such a strong high-energy tail to the spectrum.

On adding the β -ray energy of 1.792 mev to 3.204 mev, the sum of the M and X γ -ray energies, we get 4.996 mev for the difference in energy of the ground states of ThC'' and Pb^{208} . This differs by less than the estimated error from the value deduced from the α -decay energies Briggs (1936, and others), if the β -ray energy of ThC.C' is taken to be 2.2505 mev.

Level schemes which include the P γ -ray are discussed by Richardson (1948).

§ 11. THE POSSIBLE β -EXCITATION OF LEVELS OF VERY HIGH ENERGY

The graphs of $\sqrt{(N/F)}$ predicted by the level schemes can be seen to be nearly straight at their low energy ends.

The experimental curve, in contrast, shows a marked downward convexity at energies as low as $\epsilon = 1.4$.

The curvature, which is very small between $\epsilon = 3.2$ and $\epsilon = 2.1$ rises sharply near the latter point. This excess of slow β -rays may be due to one or more partial spectra of low energy. Its magnitude can be estimated by subtracting from the observed values of N the values obtained by approaching $\epsilon = 1$ along a straight line forming a tangent to the graph at about $\epsilon = 2.2$. We get surpluses of 16%, 11% and 5% in the values of N at the three points of lowest energy taken in order of increasing ϵ . These figures are compatible with a single partial spectrum with a transition probability p_{β} of about 3% and an end-point ϵ_{β} of about $2.1 mc^2$ exciting a level of thorium D at about 4.45 mev above the ground state. So large a value of p_{β} suggests that there may be several close levels and requires that the excitation is by allowed β -transitions. The selection rules for an allowed transition give the β -excited level values of spin and parity which permit quadripole γ -ray transitions to all lower energy levels also excited by allowed transitions. As there are very probably three such lower levels we expect at least three quadripole γ -rays of about 1 mev starting from a level at about 4.5 mev. Owing to the low intensities their internal conversion lines should be difficult to find, but the weak γ -ray of (1350 ± 20) kev found by Alichanow and Dzelepov (1938) may be one of these rays. Its intensity is stated by Latyshev and Kulchitsky (1941) to be 0.036 times that of the X γ -ray.

An alternative explanation of the surplus of slow β -rays is that they are generated in the aluminium leaf A by the processes of inelastic back-scattering and ejection

of secondary (knock-on) electrons in close collisions. The experiments of Flammersfeld (1939), and others, show that these effects do not become appreciable at energies of about 100 kev until the thickness of the mounting material is about 1 mg/cm². In our case with 0.13 mg/cm² they should be negligible.

A third possible explanation is that the surplus is due to an increased efficiency of the β -counter for slow rays due to their greater specific ionization. It seems unlikely that a rise in N of more than 2%, for the two lowest points can be accounted for in this way.

§ 12. CONCLUSION

The level schemes considered above all fail both to give agreement with the observed shape of the spectrum and to conform to the requirement of energy balance in the thorium C branching in a way which is compatible with the experimental value of the end-point at 1.792 mev. It seems probable that no energy levels of thorium D lying below 3.20 mev can receive more than 1 or 2% β -excitation. In order to account for this low excitation it is evident that the lower levels can only be accessible via forbidden β -transitions and that the transition to the ground level must be highly forbidden.

Owing to the presence of five strong γ -rays and at least three weak ones a great many level schemes are possible. A clue, which may, however, be a chance coincidence, is provided by the P, G, and M γ -rays, which satisfy the combination

$$(h\nu)_G + (h\nu)_M = (h\nu)_P.$$

The evidence is given in table 5.

Table 5

Author	Ellis (1932)	Surugue (1937)
Energy of G γ -ray (kev.)	276.7	276.8
Energy of M γ -ray	582.5	583.4
Sum	859.2	860.2
P β -line	771.6	772.3
Surplus	87.6	87.9

The surplus agrees fairly well with the K-ionization energy of 88.10 kev for $Z=82$. Each author's values for the energies of the β -lines have been used, but the ionization energies are those of Ingelstam tabulated by Arnoult (1939).

If we assume the spin of the ground state of 'ThC'' to be four units of $\hbar/2\pi$, the Gamow-Teller β -selection rules for an allowed transition permit the excitation of levels of ThD with spin quantum number $I=3, 4$ and 5. The distribution of convexity in the graph of $\sqrt{(N/F)}$ shows that levels with these spins are present near the high-energy end-points between 3.1 mev and 3.75 mev and that after an interval they appear again at higher energies above 4.4 mev possibly in the form of a close group of levels. The ground state of ThD has very probably $I=0$ and the spacing of the levels would be expected to get smaller at high energies of excitation.

ACKNOWLEDGMENTS

We wish to thank Sir James Chadwick for his interest and for preparing the radiothorium, Dr. J. C. P. Miller for his generous help with statistical and computing problems, and Dr. O. R. Frisch for the use of his unpublished scaling circuit.

REFERENCES

- ALICHANIAN, A. I., and ZAVELSKIY, A. S., 1937, *C.R. Acad. Sci., U.R.S.S.*, **17**, 469.
 ALICHANOW, A. I., and DZELPOW, V. P., 1938, *C.R. Acad. Sci., U.R.S.S.*, **20**, 113.
 ARNOULT, R., 1939, *Ann. Phys., Paris*, **12**, 241.
 BRIGGS, G. H., 1936, *Proc. Roy. Soc. A*, **157**, 183.
 ELLIS, C. D., 1932, *Proc. Roy. Soc. A*, **138**, 318.
 FERMI, E., 1934, *Z. Phys.*, **88**, 161.
 FLAMMERSFELD, A., 1939, *Z. Phys.*, **114**, 227.
 GURNEY, R. W., 1926, *Proc. Roy. Soc. A*, **112**, 380.
 HENDERSON, W. J., 1934, *Proc. Roy. Soc. A*, **147**, 572.
 INGELSTAM, E., 1937, *Nova Acta Reg. Soc. Sci. Uppsala*, **10**, 1.
 ITOH, J., and WATASE, Y., 1941, *Proc. Phys. Math. Soc. Japan*, **23**, 142.
 KOVARIK, A. F., and ADAMS, N. I., 1938, *Phys. Rev.*, **54**, 413.
 LATYSHEV, G. D., and KULCHITSKY, L. A., 1941, *J. Phys. Acad. Sci., U.R.S.S.*, **4**, 515.
 LAWSON, R. W., 1919, *Wien. Ber.*, **128**, 795.
 LAWSON, J. L., and TYLER, W., 1940, *Rev. Sci. Instrum.*, **11**, 6.
 LEWIS, H., and BOHM, D., 1946, *Phys. Rev.*, **69**, 129.
 MARGENAU, H., and MURPHY, G. M., 1943, *The Mathematics of Physics and Chemistry* (New York: Van Nostrand Co.).
 MARTIN, L. H., and TOWNSEND, A. A., 1939, *Proc. Roy. Soc. A*, **170**, 190.
 MEITNER, L., and PHILIPP, K., 1933, *Z. Phys.*, **80**, 277.
 NEARY, G. J., 1940, *Proc. Roy. Soc. A*, **175**, 71.
 OPPENHEIMER, F., 1936, *Proc. Camb. Phil. Soc.*, **32**, 328.
 RICHARDSON, H. O. W., 1948, *Nature, Lond.*, **161** (in press).
 SCHONLAND, B. F. J., 1925, *Proc. Roy. Soc. A*, **108**, 187.
 SURUGUE, J., 1937, *Ann. Phys., Paris*, **8**, 484.
 SZE, S. Y., 1933, *Ann. Phys., Paris*, **19**, 59.

On the Solution of Scattering and Related Problems

BY B. FERRETTI * AND M. KROOK

Department of Mathematical Physics, The University, Birmingham

* Now at University of Milan

Communicated by R. E. Peierls; MS. received 9 February 1948

ABSTRACT. A method is presented for solving scattering and eigenvalue problems. This is of use mainly in cases where the forces are of short range. The method is based on a Taylor expansion of the radial eigenfunctions about a point where the potential is already rather small. A procedure is given for determining the logarithmic derivative at this point of the solution which is "least singular" at the origin, this is in general the physically admissible solution.

§ 1. INTRODUCTION

IT is well known that for calculating scattering cross-sections one does not require the detailed solution of the relevant wave equation, but only its asymptotic form. Thus for a system with forces of finite range a , the scattering problem is effectively solved when one has determined the logarithmic derivatives of the interior radial solutions at the boundary a ; similar considerations apply also to the corresponding eigenvalue problem for the bound states. In more general cases where the forces are of short range one may solve the scattering or eigenvalue problem by subdividing the whole range $(0, \infty)$ into an "interior region" $(0, a)$ and an "exterior region" (a, ∞) . In the exterior

region, where the potential is already rather small, the solution may be obtained by a perturbation method together with the condition that this exterior solution fit smoothly on to the interior one. Here again the only information about the interior solution needed in the calculation is the value of its logarithmic derivative at the point of subdivision a .

The object of this paper is to present a method of evaluating the logarithmic derivative of the interior solution at a , without solving the equation explicitly. Usually the origin is a singular point of the radial differential equation. The customary procedure is then to expand the solution in a power series about the origin so as to satisfy a condition of regularity or finiteness there. The present method too is based on a Taylor expansion, not about the origin however, but about the boundary or subdivision point a . The condition of regularity at the origin is then equivalent to a condition on the radius of convergence of the series. It is this latter condition which permits the direct evaluation of the logarithmic derivative of the interior solution at the boundary.

§2 METHOD OF CALCULATION

In general the radial differential equation is linear, of second order and with the origin as a regular singular point :

$$\frac{d^2u}{dr^2} + f(r)\frac{du}{dr} + g(r)u = 0. \quad \dots\dots(2.1)$$

The equation may also have singular points other than $r = 0$. We are concerned with that solution of (2.1) which is regular at $r = 0$; more particularly we wish to determine its logarithmic derivative at the ordinary point $r = a$.

Let $r = 0$ be the singular point of (2.1) nearest to a ; and let d be the distance from $r = a$ to the next nearest singularity. Since a is an ordinary point of (2.1), any solution can be expanded in a Taylor series about a :

$$u(r) = \sum_{n=0}^{\infty} c_n (r-a)^n$$

where

$$c_n = \frac{1}{n!} u_n = \frac{1}{n!} \left(\frac{d^n u}{dr^n} \right)_{r=a}. \quad \dots\dots(2.2)$$

In general the radius of convergence of this series will be just a , and so by Cauchy's convergence test :

$$\lim_{n \rightarrow \infty} |c_n|^{1/n} = 1/a. \quad \dots\dots(2.3)$$

But when, and only when, $u(r)$ is that special solution which is regular at $r = 0$ the radius of convergence will exceed a and will in fact be d ; i.e.

$$\lim_{n \rightarrow \infty} |c_n|^{1/n} = 1/d. \quad \dots\dots(2.4)$$

It is immediately evident that c_n can be represented in the form :

$$c_n = F_n c_0 + G_n c_1, \quad \dots\dots(2.5)$$

where $F_0 = G_1 = 1$, $F_1 = G_0 = 0$ and the coefficients F_n , G_n , ($n = 2, 3 \dots$) are uniquely determined by the differential equation. Let $b = u'(a)/u(a)$ denote the logarithmic derivative at $r = a$ of the regular solution. We then have

$$b = -\lim_{n \rightarrow \infty} F_n/G_n. \quad \dots\dots(2.6)$$

This can be proved as follows: We are free to make either $c_0=0$ or $c_1=0$ in (2.5). Then at least one of the two corresponding functions $u(r)$ must be singular at the origin and thus lead to (2.3). Hence at least one of the relations

$$\lim_{n \rightarrow \infty} |F_n|^{1/n} = 1/a; \quad \lim_{n \rightarrow \infty} |G_n|^{1/n} = 1/a$$

is valid. For definiteness suppose the second equation holds. Then when $c_1/c_0 = b$ we have

$$\begin{aligned} 1/d &= \lim_{n \rightarrow \infty} |F_n c_0 + G_n c_1|^{1/n} \\ &= \lim_{n \rightarrow \infty} \{ |c_0|^{1/n} |G_n|^{1/n} (F_n/G_n) + b |c_1|^{1/n} \} \\ \text{and so} \quad \lim_{n \rightarrow \infty} |(F_n/G_n) + b|^{1/n} &= a/d \end{aligned} \quad \dots\dots(2.7)$$

which establishes (2.6).

Equation (2.7) indicates that the convergence to the limit in (2.6), once it "sets in", is rapid provided a/d is not too close to unity. It also provides a useful control in some applications, enabling us to decide whether or not an empirically determined convergence is real.

In general it is not feasible to attempt to find the exact limiting value b in (2.6). The main use of (2.6), (2.7) in practice is as the basis for an approximation method; and it is just this feature which we wish to emphasize in the present paper. The procedure then consists in regarding $b_n = -F_n/G_n$ for some value of n as representing b , the particular choice of n being guided by the convergence criterion and the order of accuracy required.

To illustrate the procedure we shall now discuss a number of examples: In these we shall write (cf. (2.2), (2.5)).

$$A_n u_n = A_n c_n/n! = P_n u_0 + Q_n u_1, \quad \dots\dots(2.8)$$

where A_n is some convenient numerical factor. Then

$$P_n = A_n F_n/n!, \quad Q_n = A_n G_n/n!$$

and the n th approximation to b is given by

$$b_n = -P_n/Q_n. \quad \dots\dots(2.9)$$

§ 3. EXAMPLE 1

$$\text{Consider the equation} \quad r \frac{d^2 u}{dr^2} - u = 0, \quad \dots\dots(3.1)$$

which has as solution regular at $r=0$ the function $\sqrt{r} I_1(2\sqrt{r})$, I_1 being the Bessel function of imaginary argument. Differentiating (3.1) $(n-2)$ times and substituting $r=a$ we obtain a recurrence formula for the u_n of (2.2):

$$a u_n = -(n-2) u_{n-1} + u_{n-2}. \quad \dots\dots(3.2)$$

Writing

$$a^k u_k = P_k u_0 + Q_k u_1$$

it follows that the P_k and the Q_k separately each satisfy recurrence relations of the same form (3.2):

$$\begin{aligned} P_n &= -(n-2)P_{n-1} + aP_{n-2}, \\ Q_n &= -(n-2)Q_{n-1} + aQ_{n-2}, \quad \} \quad \dots\dots(3.3) \\ P_0 &= 1; \quad Q_0 = 0 = P_1; \quad Q_1 = a. \end{aligned}$$

Repeated application of (3.3) then leads to :

$$\begin{array}{ll}
 P_2 = a & Q_2 = 0 \\
 P_3 = -a & Q_3 = a^2 \\
 P_4 = 2a + a^2 & Q_4 = -2a^2 \\
 P_5 = -(6a + 4a^2) & Q_5 = 6a^2 + a^3 \\
 P_6 = 24a + 18a^2 + a^3 & Q_6 = -(24a^2 + 6a^3) \\
 P_7 = -(120a + 96a^2 + 9a^3) & Q_7 = 120a^2 + 36a^3 + a^4 \\
 P_8 = 720a + 600a^2 + 72a^3 + a^4 & Q_8 = -(720a^2 + 240a^3 + 12a^4) \\
 P_9 = -(5040a + 4320a^2 + 600a^3 + 16a^4) & Q_9 = 5040a^2 + 1800a^3 + 120a^4 + a^5.
 \end{array}$$

The n th approximation, b_n , to $u'(a)/u(a)$ is given by (2.9). To indicate the order of accuracy attainable and the mode of convergence, successive values of b_n for three different values of a are listed in table 1. (The entries for $n = \infty$ were obtained by use of tables of Bessel functions).

Table 1

n	$b_n (a=1)$	$b_n (a=4)$	$b_n (a=9)$
3	1.000000	0.250000	0.111111
4	1.500000	0.750000	0.611111
5	1.428571	0.550000	0.311111
6	1.433333	0.583333	0.380342
7	1.433121	0.578571	0.362540
8	1.433127	0.579060	0.365300
9	1.433127	0.579022	0.365300
∞	1.433127	0.579024	0.365353

Even for comparatively large values of the argument accurate values of b can be obtained rapidly. To obtain comparable accuracy with the use of a Taylor expansion about the origin requires the retention of a large number of terms and more laborious computations.

It is interesting to remark that the procedure may be extended to obtain representation of the function $u(a) = \sqrt{a} I_1(2\sqrt{a})$ over a fairly wide range in terms of elementary functions. Regarding $b_8 = -P_8/Q_8$ as representing $u'(a)/u(a)$ with sufficient accuracy in a range $(0, A)$, an elementary integration leads to the result :

$$u(a) = ae^{a/12} \left(\frac{a^2 + 20a + 60}{60} \right)^{5/3} \left[\frac{(10 - 2\sqrt{10})a + 60}{(10 + 2\sqrt{10})a + 60} \right]^{25/12 \sqrt{10}}.$$

Here the constant of integration has been fixed by using the known behaviour of $u(a)$ in the neighbourhood of the origin. The accuracy of the formula (3.4) decreases gradually as a increases. At $a=9$ it is in error by less than 0.01 %.

§ 4. EXAMPLE 2

Consider the equation $r(r+1) \frac{d^2 u}{dr^2} - u = 0$ (4.1)

and suppose that the logarithmic derivative of the regular solution is required for $r=2$. Here the direct method of series expansion about the origin breaks down on account of the singularity at $r=-1$; the series would have radius of convergence unity. (Of course an expansion about the origin can still be used in conjunction with a bilinear substitution, e.g. one which transforms -1 into ∞). It is here that the method of this paper can be used with particular advantage.

$$\text{Writing} \quad (a^2 + a)^n u_n = P_n u_0 + Q_n u_1 \quad \dots\dots(4.2)$$

we obtain the recurrence formula :

$$\left. \begin{aligned} P_n &= -(n-2)(2a+1)P_{n-1} - [(n-2)(n-3)-1](a^2+a)P_{n-2} \\ Q_n &= -(n-2)(2a+1)Q_{n-1} - [(n-2)(n-3)-1](a^2+a)Q_{n-2} \end{aligned} \right\} \quad \dots\dots(4.3)$$

with $P_0=1$, $Q_0=P_1=0$, $Q_1=(a^2+a)$.

Proceeding as in § 3, successive approximations to u_1/u_0 (with $a=2$) are given by $b_n = -P_n/Q_n$:

$$\begin{array}{llll} b_3=0.8333 & b_4=0.7333 & b_5=0.7083 & b_6=0.6988 \\ b_7=0.6945 & b_8=0.6923 & b_9=0.6911 & b_{10}=0.6903. \end{array}$$

Here the differences between successive terms b_k , ($k>4$), agree quite well with the convergence criterion $|b-b_n| \sim K(2/3)^n$ with K depending only slightly on n . This fact may be used to obtain a more accurate value for b :

$$b'_n = b_n - K(2/3)^n = b_n - 2(b_{n-1} - b_n).$$

The corresponding values of b'_n are $b'_8=0.6879$; $b'_9=0.6887$; $b'_{10}=0.6887$.

§ 5. EXAMPLE 3

The differential equations

$$\frac{d^2 u}{dx^2} + \left[k^2 - \frac{l(l+1)}{x^2} + B \frac{e^{-x}}{x} \right] u = 0 \quad \dots\dots(5.1)$$

($l=0, 1, 2, \dots$) occur in the theory of scattering with a central "meson potential". We consider the case $l=0$ and write

$$u = xV \quad \dots\dots(5.2)$$

$V(x)$ must then be finite at $x=0$. By repeated differentiation of the differential equation for $V(x)$ we obtain for $V_n = V^{(n)}(a)$ the recurrence formula :

$$aV_{n+2} + (n+2)V_{n+1} + k^2[aV_n + nV_{n-1}] = -Be^{-a} \sum_{k=0}^n (-1)^k \binom{n}{k} V_{n-k}. \quad \dots\dots(5.3)$$

Writing $a^m V_m = P_m V_0 + Q_m V_1$ with $P_0=1$, $Q_0=P_1=0$, $Q_1=a$, we obtain recurrence formulae for P_m , Q_m separately :

$$P_{n+2} = -(n+2)P_{n+1} - k^2 a^2 (P_n + nP_{n-1}) - aBe^{-a} \sum_{k=0}^n (-1)^k \binom{n}{k} P_n k a^k$$

with an identical formula for the Q_m . Taking the case $k=0.8$, $B=1.5$ and $l=0$, for which the asymptotic phase has been obtained by numerical integration (Sachs and Goepfert-Mayer 1938), the following values of $b_n = -P_n/Q_n$ are obtained with $a=2$:

$$\begin{array}{llll} b_5 = -1.161 & b_6 = -1.284 & b_7 = -1.348 & b_8 = -1.375 \\ b_9 = -1.389 & b_{10} = -1.396 & b_{11} = -1.399. & \end{array}$$

The differences* between successive b_n , ($n>4$), behave regularly and provide an improved extrapolated value for $b = V'(2)/V(2)$: $b = -1.401$, and so

$$u'(2)/u(2) = -0.901.$$

* In most cases it is possible to obtain an improved value in this way by extrapolation using the differences between successive values of b_n .

This provides an accurate value for the phase shift $\bar{\eta}$ of the exact wave function at $x=2$: $\bar{\eta}=0.8154$.

The asymptotic phase η_0 satisfies the exact equation *

$$k \sin(\eta_0 - \bar{\eta}) = B \int_2^{\infty} \frac{e^{-x}}{x} u(x) \sin(kx + \bar{\eta}) dx. \quad \dots\dots(5.4)$$

Since the potential is already rather small for $x=2$, a fairly close approximation to η_0 can be obtained by evaluating the integral in (5.4) with $\sin(kx + \bar{\eta})$ substituted for $u(x)$. A straightforward numerical evaluation of the integral then leads to $\sin(\eta_0 - \bar{\eta}) = 0.0217$ and so $\eta_0 = 0.8371$. This is in good agreement with the value $\eta_0 = 0.83706$ obtained by Sachs and Goepfert-Mayer (1938).

For higher values of l the procedure converges more rapidly than for $l=0$; the larger the value of l the more rapid is the convergence. This feature is not peculiar to the particular potential function in (5.1) but holds in general for arbitrary forms of the potential. It can easily be understood in terms of the relative degree of singularity of the regular and irregular solutions which increases with l .

An alternative method of performing the calculation is to write the solution of (5.1) in the form

$$u(x) = f(x)\psi_l(kx) + g(x)\chi_l(kx)$$

where $\psi_l(t) = \sqrt{(\pi t/2)} J_{l+1/2}(t)$, $\chi_l(t) = \sqrt{(\pi t/2)} J_{-l-1/2}(t)$. $J_{l+1/2}$ and $J_{-l-1/2}$ are Bessel functions of the first kind.

Imposing the condition

$$f'(x)\psi_l(kx) + g'(x)\chi_l(kx) = 0$$

we obtain a pair of first order equations for f, g :

$$\left. \begin{aligned} f'(x) &= -\frac{(-1)^l}{k} v(x) u(x) \chi_l(kx), \\ g'(x) &= +\frac{(-1)^l}{k} v(x) u(x) \psi_l(kx). \end{aligned} \right\} \quad \dots\dots(5.5)$$

The condition $f^{(n)}(a) = 0$ then leads to an approximation for $g(a)/f(a)$ and so to an approximation for the phase shift at $x=a$. The repeated differentiations of (5.5) are more complicated than in the above direct method, but f, g are more slowly varying functions than u and the procedure would thus converge more rapidly.

§ 6. APPLICATION TO TENSOR FORCES

In the discussion of neutron-proton scattering with tensor forces one encounters the pair of simultaneous equations (cf. Rarita and Schwinger 1941),

$$\frac{d^2 u}{dx^2} + [k^2 + v_1(x)]u = v_3(x)w; \quad \frac{d^2 w}{dx^2} + \left[k^2 - \frac{6}{x^2} + v_2(x) \right]w = v_3(x)u \quad \dots\dots(6.1)$$

associated with the triplet state of unit angular momentum and even parity. v_1, v_2, v_3 are specified potential functions. The physical solutions are required to satisfy, amongst other conditions, that of vanishing at least like x at the origin. For simplicity we shall consider the case of a rectangular potential; v_1, v_2, v_3 are then constant in $(0, 1)$ and vanish for $x > 1$.

* (5.4) is easily obtained e.g. by a slight modification of the argument leading to equation (6.6) in Bethe and Bacher (1936).

An examination of (6.1) in the neighbourhood of the origin shows that the physically admissible solutions are of the form :

$$\begin{aligned} u(x) &= \sum_{n=0}^{\infty} A_n x^{n+1} + \ln x \sum_{n=0}^{\infty} C_n x^{n+5}, \\ w(x) &= \sum_{n=0}^{\infty} B_n x^{n+3} + \ln x \sum_{n=0}^{\infty} D_n x^{n+3}. \end{aligned} \quad (6.2)$$

The constants A_0, B_0 can be assigned arbitrary values independently. There are thus exactly two independent solutions which vanish as required at $x=0$. On account of the linearity of (6.1) any linear combination of solutions is again a solution. A fundamental system of solutions

$$\left. \begin{aligned} u &= \alpha(x) \\ w &= \beta(x) \end{aligned} \right\} \quad \left. \begin{aligned} u &= \phi(x) \\ w &= \psi(x) \end{aligned} \right\} \quad \dots\dots (6.3)$$

can then be chosen so as to satisfy two arbitrary independent pairs of conditions at $x=1$. In particular, there exists one such system with the property

$$\left. \begin{aligned} \alpha(1) &= 1 \\ \beta(1) &= 0 \end{aligned} \right\} \quad \left. \begin{aligned} \phi(1) &= 0 \\ \psi(1) &= 1 \end{aligned} \right\} \quad \dots\dots (6.4)$$

Once the derivatives $\alpha'(1), \beta'(1), \phi'(1)$ and $\psi'(1)$ are known for the system (6.3) satisfying (6.4) (or any other set of conditions), the scattering problem is, in effect, solved. It only remains to calculate the asymptotic phases by fitting appropriate exterior solutions (expressible in terms of Bessel functions) on to the interior solutions at $x=1$. A knowledge of the above first derivatives of the interior solution is sufficient for this purpose.

The discussion in the preceding sections has been based on the translation of a regularity condition at the origin into a condition on the radius of convergence of a power series. This point of view is, however, unnecessarily restricted. The selection of the regular solution is a special case of a more general property of the procedure. It is in fact applicable to problems in which the required solution is not regular at the origin. In this case the method selects the "least singular" solution consistent with the specified data*. A proof of this, in a form sufficiently general for our purposes, is given in the Appendix.

Now, the physical solutions of type (6.2), although not regular at the origin, have the property of being "least singular" in the sense of the Appendix. The remaining solutions become infinite at $x=0$. Our method can thus be extended to apply to this problem.

We introduce new dependent variables U, W :

$$u = xU, \quad w = x^2W. \quad \dots\dots (6.5)$$

Then repeated differentiation of the differential equations for U, W , leads to recurrence relations for $U_m - U^{(m)}(1)$ and $W_m - W^{(m)}(1)$. Writing

$$\begin{aligned} X_n &= U_n + nU_{n-1}, \\ Y_n &= W_n + 2nW_{n-1} + n(n-1)W_{n-2}, \\ Z_n &= (k^2 + v_1)X_n - v_3Y_n, \\ T_n &= (k^2 + v_2)Y_n - v_3X_n, \end{aligned} \quad (6.6)$$

* We are indebted to Professor Peierls for this suggestion.

we have
$$\left. \begin{aligned} U_{n+2} &= -(n+2)U_{n+1} - Z_n, \\ W_{n+2} &= -(2n+4)W_{n+1} - (n-1)(n+4)W_n - T_n. \end{aligned} \right\} \dots\dots (6.7)$$

(The relations $X_{n+2} = -Z_n$; $Y_{n+2} = 6W_n - T_n$ are useful in application of (6.6), (6.7)).

We now write

$$\left. \begin{aligned} U_m &= P_m U_0 + Q_m U_1 + R_m W_0 + S_m W_1, \\ W_m &= K_m U_0 + L_m U_1 + M_m W_0 + N_m W_1. \end{aligned} \right\} \dots\dots (6.8)$$

Then

$$\left. \begin{aligned} P_0 &= Q_1 = M_0 = N_1 = 1. \\ \text{All other } P_m, \dots, N_m &\text{ with } m=0, 1 \text{ zero.} \end{aligned} \right\} \dots\dots (6.9)$$

It is evident that the pair of coefficients (P_m, K_m) , $(m=0, 1, 2 \dots)$ by themselves satisfy the recurrence relations (6.6), (6.7) with P_m replacing U_m and K_m replacing W_m ; the same is true for the remaining pairs (Q_m, L_m) , (R_m, M_m) , (S_m, N_m) . This simplifies the calculation of successive U_m, W_m .

The procedure is now to set both $U_m=0$ and $W_m=0$ in (6.8). Putting $U_0=1, W_0=0$ then leads to the equations :

$$Q_m U_1 + S_m W_1 = -P_m; \quad L_m U_1 + N_m W_1 = -K_m. \quad \dots\dots (6.10)$$

The values of U_1, W_1 thus obtained provide the required first derivatives

$$\alpha'(1) = 1 + U_1, \quad \beta'(1) = W_1$$

of one solution of the fundamental system (6.3, 6.4). $\phi'(1), \psi'(1)$ are obtained in the same way by solution of :

$$Q_m U_1 + S_m W_1 = -R_m; \quad L_m U_1 + N_m W_1 = -M_m$$

together with

$$\phi'(1) = U_1; \quad \psi'(1) = W_1 + 2.$$

To illustrate the mode of convergence of the procedure, successive approximations to $(\alpha'(1), \beta'(1); \phi'(1), \psi'(1))$ are listed in table 2. As our object in the present paper is exposition of method rather than actual calculation of cross-sections the calculations for table 2 were performed with the simple values

$$k^2 + v_1 = 3, \quad k^2 + v_2 = -1, \quad v_3 = -6.$$

Table 2

n	$\alpha'(1)$	$\beta'(1)$	$\phi'(1)$	$\psi'(1)$
3	-7.4615	-3.2308	+0.4615	+3.2308
4	+0.5000	-1.3500	-2.8750	+2.5625
5	-0.3421	-1.3953	-1.1245	+3.1031
6	-0.8654	-1.5320	-1.7084	+2.8668
7	-1.0189	-1.5711	-1.5093	+2.9178
8	-0.8721	-1.5316	-1.5381	+2.9109
9	-0.8634	-1.5287	-1.5260	+2.9141
10	-0.8694	-1.5300	-1.5305	+2.9131
11	-0.8690	-1.5297	-1.5292	+2.9134
12	-0.8687	-1.5293	-1.5291	+2.9135

APPENDIX

Let $f(z)$ and $g(z)$ be analytic functions having no singular point other than $z=0$ in some circle $|z-a|=b$, ($b>|a|$). Without loss of generality, a may be assumed real and positive. If either f or g or both are many-valued with the

origin as a branch point, a cut is made along the negative real axis from 0 to ∞ ; we then restrict our considerations to a single branch of f and g .

Suppose, further, that in the cut plane: (a) there exists a real number σ , ($0 < \sigma < 1$), such that

$$z^{1-\sigma}f(z) \rightarrow 0 \text{ as } z \rightarrow 0, \quad \dots\dots(7.1)$$

(b) $g(z)$ can be expressed as :

$$g(z) = \phi(z) + \psi(z), \quad \dots\dots(7.2)$$

where $\phi(z)$ satisfies the same conditions as $f(z)$; the "more singular" part $\psi(z)$ will be assumed to have either of the forms :

$$\psi(z) = \sum_{k=1}^{\nu} \frac{A_k}{z^{\lambda_k}}, \quad \text{. (I)}$$

$$\psi(z) = \ln z \sum_{k=1}^{\nu} \frac{B_k}{z^{\lambda_k}}, \quad \text{. (II)}$$

where $\lambda_1 > \lambda_2 > \dots > \lambda_{\nu} \geq 1$.

We consider the Taylor series

$$f(x) = \sum_{n=0}^{\infty} f_n(x-a)^n,$$

$$g(x) = \sum_{n=0}^{\infty} g_n(x-a)^n = \sum_{n=0}^{\infty} (\phi_n + \psi_n)(x-a)^n$$

and shall prove that $\lim f_n/g_n = 0$.

Let $h < b-a$ and let \mathcal{C}_δ denote the contour $\Gamma + \mathcal{L}_\delta + \gamma_\delta + \mathcal{L}'_\delta$ in the cut plane. Γ is the circle with centre a and radius $a+h$, γ_δ the circle $|z|=\delta$; \mathcal{L}_δ and \mathcal{L}'_δ are straight-line paths between $-h$ and $-\delta$ along the upper and lower lips of the cut respectively (see figure). Then

$$\begin{aligned} f_n &= \frac{1}{2\pi i} \int_{\mathcal{C}_\delta} \frac{f(z)}{(z-a)^{n+1}} dz \\ &= \frac{1}{2\pi i} \left\{ \int_{\Gamma} + \int_{\mathcal{L}_\delta} + \int_{\gamma_\delta} + \int_{\mathcal{L}'_\delta} \right\} \frac{f(z) \cdot dz}{(z-a)^{n+1}}. \end{aligned}$$

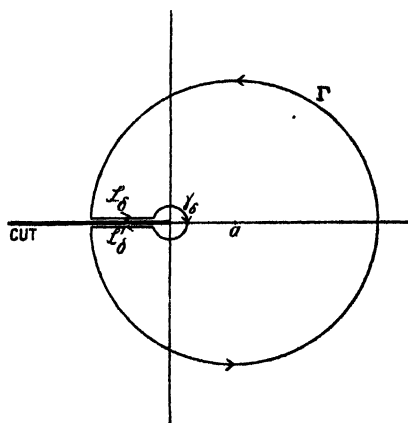
Let $M_1(h) = \text{maximum of } |f(z)| \text{ on } \Gamma$,

$m(h) = \text{maximum of } |z^{1-\sigma}f(z)| \text{ on the two lines } -h \leq z < 0$,

$M_2(\delta) = \text{maximum of } |f(z)| \text{ on } \gamma_\delta$.

Then

$$\begin{aligned} \left| \frac{1}{2\pi i} \int_{\mathcal{L}_\delta} \frac{f(z) dz}{(z-a)^{n+1}} \right| &= \frac{1}{2\pi} \left| \int_{\mathcal{L}_\delta} \frac{z^{1-\sigma}f(z) dz}{z^{1-\sigma}(z-a)^{n+1}} \right| \\ &\leq \frac{1}{2\pi} \frac{m(h)}{(a+\delta)^{n+1}} \int_{\mathcal{L}_\delta} \left| \frac{dz}{z^{1-\sigma}} \right| \\ &\leq \frac{h^\sigma m(h)}{2\pi\sigma(a+\delta)^{n+1}}. \end{aligned}$$



Similarly for the integral along \mathcal{L}'_0 . Hence

$$|f_n| \leq \frac{M_1(h)}{(a+h)^n} + \frac{2h^a m(h)}{2\pi\sigma(a+\delta)^{n+1}} + \frac{\delta M_2(\delta)}{(a-\delta)^{n+1}}.$$

By (7.1), $\delta M_2(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. Thus, proceeding to the limit $\delta \rightarrow 0$:

$$|f_n| \leq \frac{1}{a^n} \left\{ M_1(h) \left(1 + \frac{h}{a}\right)^{-n} + \frac{h^a \cdot m(h)}{\sigma\pi a} \right\}$$

Now, given $\epsilon > 0$, however small, we can choose h so that

$$\frac{h^a m(h)}{\pi\sigma a} < \frac{\epsilon}{2}.$$

Then we can find a number N_0 such that for all $n > N_0$

$$M_1(h)(1+h/a)^{-n} < \epsilon/2.$$

Hence, for all $n > N_0$,

$$|f_n| < \epsilon/a^n. \quad \dots\dots(7.3)$$

On the other hand, by direct differentiation of $\psi(z)$ in the forms (I), (II) we have

$$(-1)^n \frac{A_1}{a^{\lambda_1+1+n}} \binom{n+\lambda_1-1}{n} + \dots \quad \dots\dots(I)$$

$$\psi_n \equiv \frac{\psi^{(n)}(a)}{n!} = \left\{ \begin{aligned} &(-1)^n \frac{B_1}{a^{\lambda_1+1+n}} \left\{ \binom{n+\lambda_1-1}{n} \ln a - \sum_{k=1}^n \frac{1}{k} \binom{n+\lambda_1-k}{n-k} \right\} \\ &+ \dots \end{aligned} \right. \quad \dots\dots(II)$$

where the dots denote terms which are of lower order as $n \rightarrow \infty$. In both cases the coefficient of $a^{-(\lambda_1+n)}$ is at least non-zero: in general it becomes infinite as $n \rightarrow \infty$ (except for $\lambda_1 = 1$ in case (I)). Thus, for both (I) and (II), there exists a number N_1 such that, for all $n > N_1$

$$|\psi_n| > M_3/a^{\lambda_1+n}, \quad \dots\dots(7.4)$$

where M_3 is a constant independent of n .

Thus, combining (7.3), (7.4), it follows that for all $n > \text{Max}(N_0, N_1)$, $|f_n/\psi_n| < a^{\lambda_1} \epsilon/M_3$, and so $\lim_{n \rightarrow \infty} f_n/\psi_n = 0$.

Since $\phi_n/\psi_n \rightarrow 0$ as $n \rightarrow \infty$ (ϕ_n behaves "at worst" like f), we have $\lim_{n \rightarrow \infty} f_n/g_n = 0$.

If now $f(x)$, $g(x)$ form a fundamental system of solutions for a given second order differential equation, the general solution is of the form

$$u = \sum_{n=0}^{\infty} u_n (x-a)^n = Af(x) + Bg(x). \quad \text{Then } u_n = Af_n + Bg_n.$$

Our procedure is to set $u_n = 0$. This gives $B/A = -f_n/g_n$ which $\rightarrow 0$ as $n \rightarrow \infty$. In this way the "least singular" solution $f(x)$ is selected.

The proof can easily be extended to the case of a differential equation of order k , with $(k-1)$ conditions specified at $x=a$; the method will again select the "least singular" solution consistent with the specified data.

ACKNOWLEDGMENTS

We would like to record our thanks to Professor Peierls for his interest in this work and for a number of stimulating discussions, and to Mr. T. H. R. Skyrme for useful suggestions and criticism in connection with the mathematical theorem of the Appendix.

REFERENCES

- BETHE, H. A., and BACHER, R. F., 1936, *Rev. Mod. Phys.*, **8**, 82.
 RARITA, W., and SCHWINGER, J., 1941, *Phys. Rev.*, **59**, 434.
 SACHS, R. G., and GOEPFERT-MAYER, M., 1938, *Phys. Rev.*, **53**, 991.

Characteristics of Radio Echoes from Meteor Trails: I. The Intensity of the Radio Reflections and Electron Density in the Trails

BY A. C. B. LOVELL AND J. A. CLEGG

Physical Laboratories, University of Manchester

Communicated by P. M. S. Blackett ; MS. received 26 January 1948

ABSTRACT. Recent observations of radio echoes reflected from meteor trails have given a large amount of information about the electron densities in the trails and the physical processes occurring in the atmosphere after the passage of the meteor. In this paper, which is Part I of three papers dealing with the work, formulae are derived for the intensity of the radio wave scattered from a meteor trail on the assumption that the electrons are created in a long narrow column, of diameter small compared with the wavelength of the radio wave. Experimental work is described which shows that the predicted variation of received power with wavelength is correct for the wavelength range 4.2 m. to 8.3 m., and according to preliminary results down to 1.4 m. The formulae can then be used to measure the electron density in the trails of meteors. Of particular interest are the measurements for meteors which are also observed visually. The results show that the density in the trail of a 5th magnitude meteor (on the limit of naked eye visibility) is approximately 2×10^{10} electrons per cm. path. Brighter meteors (\sim magnitude +1) produce 10^{12} electrons per cm. path. These results are in very good agreement with contemporary theoretical calculations. The theory is compared with earlier calculations made by Eckersley and Pierce.

§ 1. INTRODUCTION

WHEN a meteor enters the earth's atmosphere it creates a trail of electrons at a height of about 100 km. above the earth's surface. The trail is dense enough to reflect energy from an incident radio wave and under suitable conditions a short-lived response can be obtained on the receiver. These transient radio responses have been studied since 1929 (see Lovell (1948) for a review of this work) and recently the technique has been developed for the study of the activity of meteor showers and the delineation of meteor radiants. (Appleton and Naismith 1947, Hey and Stewart 1946, 1947, Prentice, Lovell and Banwell 1947, Lovell, Banwell and Clegg 1947). The present papers are concerned with the mechanism of reflection of the radio waves and the physical processes occurring in the atmosphere after the creation of the meteor trail. In Part I formulae are derived for the intensity of a radio echo received from a column of electrons in the atmosphere; from these can be calculated the density of electrons in meteor trails. Part II will deal with the distribution of electron densities in the trails of the shower meteors and with the distribution of meteoric mass. In Part III the factors which determine the duration of the transient echoes and the behaviour of the trails after formation will be discussed.

§ 2. CALCULATION OF THE INTENSITY OF THE RADIO REFLECTION FROM AN ELECTRON CLUSTER

Blackett and Lovell (1941) have calculated the intensity of a radio wave scattered from a long narrow column of ionization in the atmosphere, assuming that the electrons scatter coherently. These calculations will now be extended to include the parameters of the radio apparatus and will be applied to the particular case of meteor trails.

Consider a cluster of N electrons at a distance R cm. from a radio transmitter and receiver working on a wavelength of λ cm. If the size of the cluster is small compared with λ and if the electrons reach their full velocity under the influence of the impressed E.M.F. and are not impeded by collisions, then the cross section for scattering by the N electrons will be $(8\pi/3)(e^2/mc^2)^2 N^2 \text{ cm}^2$, where e is the charge, m the mass of the electron and c the velocity of light.

If the peak power in the pulse of the radio transmitter is P watts, and if the aerial system of the transmitter has a power gain G' over a source radiating uniformly through 4π , then the power density at a distance R cm. will be $PG'/4\pi R^2 \text{ watts/cm}^2$. Hence the amount of energy scattered by the electron cluster will be $\frac{8}{3}\pi \left(\frac{e^2}{mc^2}\right)^2 N^2 \frac{PG'}{4\pi R^2} \text{ watts}$.

If the receiver is situated in the same place as the transmitter, the power density of this scattered radiation when it returns to the receiving aerial will be $\frac{8}{3}\pi \left(\frac{e^2}{mc^2}\right)^2 N^2 \frac{PG'}{4\pi R^2} \frac{1.5}{4\pi R^2} \text{ watts/cm}^2$. (The factor 1.5 is introduced to allow for the fact that the electrons scatter as Hertzian dipoles and not uniformly through 4π).

The effective collecting area of the receiving aerial will be $G_0'\lambda^2/4\pi \text{ cm}^2$ where G_0' is its power gain over a spherical source. Hence the power which it delivers to a matched load when situated in a field of mean power density σ will be $G_0'\lambda^2\sigma/8\pi$. Thus in the case under consideration the amount of energy delivered to the receiver input after scattering from the cluster of N electrons will be

$$\epsilon = \frac{8}{3}\pi \left(\frac{e^2}{mc^2}\right)^2 N^2 \frac{1.5PG'}{16\pi^2 R^4} \frac{G_0'\lambda^2}{8\pi} \text{ watts.} \quad \dots\dots(1)$$

For comparison with experiment it is convenient to take the gain of the transmitting and receiving aerials as the same, and to express the gain in terms of that of a half-wave dipole so that $G' = G_0' = 1.64G$, where G is the power gain of the aerial system over a half-wave dipole. Then (1) becomes

$$\epsilon = \frac{8}{3}\pi \left(\frac{e^2}{mc^2}\right)^2 N^2 \frac{PG^2\lambda^2}{32\pi^3 R^4} \text{ watts.} \quad \dots\dots(2)$$

In terms of the voltage amplitude V at the receiver input

$$V^2 = 2r\epsilon, \quad \dots\dots(3)$$

where r is the input resistance of the receiver. Or for constant P

$$V = k \lambda N / R^2, \quad \dots\dots(4)$$

where the constant k is given by $k^2 = \frac{8}{3}\pi \left(\frac{e^2}{mc^2}\right)^2 \frac{PG^2}{32\pi^3} 2r$.

§ 3. APPLICATION TO THE CASE OF A METEOR TRAIL

If it is assumed that the meteor trail consists of a long narrow column of ionization whose diameter is small compared with λ , the appropriate number of electrons N to be included in (2) can be calculated by using ordinary optical diffraction theory. The voltage amplitude returned to the receiver from a small element of track at an angle θ to the perpendicular from the receiver to the trail is given by $dV = k\lambda\alpha d\theta/R$, where α is the number of electrons/cm. path in the trail and k is the constant occurring in equation (4) above.

The voltage amplitude from the portion of the trail $\theta=0$ to $\theta=\theta_1$ is then given by Fresnel's integral in the form

$$V_{\theta_1} = \frac{k\alpha}{2} \left(\frac{\lambda}{R} \right)^{3/2} \int_0^{\nabla} \cos \frac{\pi}{2} \nabla^2 d\nabla + \frac{ik\alpha}{2} \left(\frac{\lambda}{R} \right)^{3/2} \int_0^{\nabla} \sin \frac{\pi}{2} \nabla^2 d\nabla,$$

where $\nabla = 2\theta\sqrt{(R/\lambda)}$.

The voltage amplitude at the receiver due to the whole track is then $V = (1/\sqrt{2})k\alpha(\lambda/R)^{3/2}$. Hence by comparison with (4) the appropriate number of electrons N is given by

$$N = \alpha\sqrt{(\lambda R/2)}. \quad \dots\dots(5)$$

In this calculation the phase correction is applied twice because of the curvature of the incident and reflected wave surfaces. The wavelength is then effectively reduced to $\lambda/2$, giving equation (5) instead of the more familiar form $N = \alpha\sqrt{(\lambda R)}$ used by Blackett and Lovell (1941).

Hence from (2) the amount of energy at the receiver input due to scattering from a meteor trail with an electron density of α electrons/cm., will be

$$\epsilon = \frac{8}{3} \pi \left(\frac{e^2}{mc^2} \right)^2 \left(\alpha^2 \frac{\lambda R}{2} \right) \left(\frac{PG^2\lambda^2}{32\pi^3 R^4} \right) = \alpha^2 \frac{PG^2\lambda^3}{24\pi^2 R^3} \left(\frac{e^2}{mc^2} \right)^2 \text{ watts.} \quad \dots\dots(6)$$

$$\text{Alternatively} \quad \alpha = \sqrt{24\pi} \left(\frac{mc^2}{e^2} \right) \frac{1}{G} \sqrt{\left(\frac{\epsilon R^3}{P\lambda^3} \right)}. \quad \dots\dots(7)$$

§ 4. THE EFFECT OF ELECTRON COLLISIONS

The above calculations assume a cross section for scattering of

$$(8\pi/3)(e^2/mc^2)^2 N^2.$$

This is correct only if the electrons are free, and it is necessary to examine the assumption, since an electron may collide with a neutral atom or molecule before it has attained its full velocity under the influence of the impressed E.M.F. This effect was neglected in the calculations of Blackett and Lovell (1941) and it will be considered here for the particular case of meteor trails.

Let ν_c be the collisional frequency of electrons with neutral atoms and ν_0 the radio wave frequency. Then the cross section for scattering is effectively reduced to

$$(8\pi/3)(e^2/mc^2)^2 N^2 (1 + (\nu_c/\pi\nu_0)^2)^{-1}. \quad \dots\dots(8)$$

Except at high altitudes this term has a considerable influence in any practicable experiment. For example at 20 km. altitude in the atmosphere an electron will make approximately 10^{11} collisions/sec. with atmospheric molecules. Now, even with a wavelength as low as 10 cm. ($\nu_0 = 3 \times 10^9$ /sec.), $(\nu_c/\pi\nu_0)^2 \simeq 100$; and for wavelengths in the region of 4 m. it rises to about 2.5×10^4 .

In the particular case of meteors the radio echoes are observed to originate at a height of approximately 100 km. At this altitude $\nu_c \simeq 3 \times 10^5$ /sec. and except for long wavelengths (e.g. $\lambda > 100$ m.) the effect of damping can be neglected. The assumption made in deducing equation (6), that the electrons are free, is therefore justified.

§ 5. EXPERIMENTAL TEST OF THE THEORY

An attempt to verify equation (6) experimentally as regards the predicted variation of ϵ with λ has been made by observing the same meteor trail simultaneously on different wavelengths. The essential features of the apparatus

have been described elsewhere (Prentice, Lovell and Banwell 1947; Lovell, Banwell and Clegg 1947). The wavelengths used were 8.3 m., 6.5 m., 4.2 m., and 1.4 m., and the main characteristics of the apparatus are summarized in table 1. A preliminary account of the measurements on 8.3 m., 6.5 m. and 4.2 m. has been given by Lovell (1947).

Table 1

λ (m.)	(a)	(b)	(c)	(d)
8.3	150	15	50	4.2
6.5	50	15	50	4.6
4.2	50	8	50	2.2
1.4	50	2	180	2.2

(a) peak transmitter power P (kw.); (b) pulse width (μ sec.); (c) pulse recurrence frequency; (d) receiver sensitivity: voltage input for twice signal/noise (μ v.).

For each equipment the aerial systems consisted of half-wave dipoles mounted at equivalent heights above the ground such that in equation (6) G can be treated as constant. Observations of the transient echoes were made visually on a cathode-ray tube display. The quantity measured is the amplitude of the echo in terms of the normal receiver noise level. This is converted into equivalent voltage input at the aerial by appropriate calibration of the receiver. Thus if simultaneous echoes are observed from the same meteor trail on wavelengths λ_1, λ_2 of amplitude V_1, V_2 equation (6) predicts that $V_1/V_2 = (\epsilon_1/\epsilon_2)^{1/2} = (\lambda_1/\lambda_2)^{1/5}$ since R, α are constants for the same trail, G is assumed to be the same and the equivalent voltage inputs are corrected for any variations in P by assuming $V_1/V_2 = (P_1/P_2)^{1/4}$.

In the actual reduction of observations the exponent n in the expression

$$V_1/V_2 = (\lambda_1/\lambda_2)^n \quad \dots\dots (9)$$

has been calculated for each pair of simultaneous observations for comparison with the theoretical value of $n = 1.5$.

(i) *Simultaneous Observations on 72 Mc/s. (4.2 m.) and 36 Mc/s. (8.3 m.)**

Simultaneous observations of the radio echoes from the same meteor trail on wavelengths of 4.2 m. and 8.3 m. using apparatus with the characteristics specified in table 1 have been made in 1947 as follows: 12-27 August, 104 observations on the Perseid radiant (21 observations rejected); 23-24 September, 158 observations on an unknown radiant (17 rejected); 9-10 October, 59 observations on an unknown radiant (2 rejected).

For each of the 321 pairs of observations the exponent n has been calculated from equation (9) as described above. The distribution curve for these measurements is plotted in figure 1 and gives a mean value of 1.4 for n with a standard deviation of 0.5.

* *Note added in proof* Equation (6) applies to the case of scattering from a meteor trail which crosses through the aerial beam in such a way that the foot of the perpendicular from the station to the trail lies within the beam. This is the only condition under which the main radio echo can be observed on wavelengths shorter than 6.5 m. On longer wavelengths some experimental evidence exists that the trails lose this critical aspect sensitivity (Millman, McKinley and Burland, *Nature, Lond.*, 1948, 161, 278, Lovell 1948). The measurements on 8.3 m. in this paper apply to the case of perpendicular reflection, since the only echoes selected are those appearing simultaneously on the shorter wavelength, which implies that the aspect condition is satisfied.

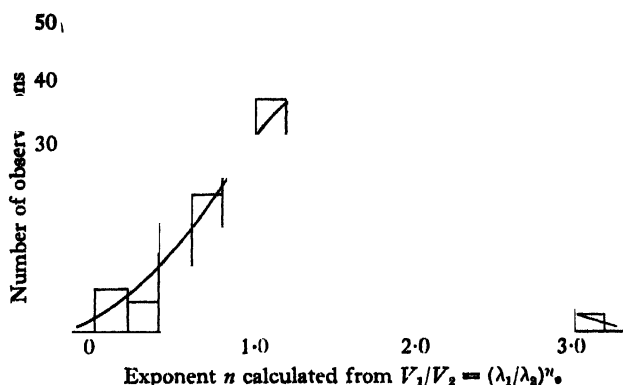


Figure 1. Distribution curve for simultaneous observations on $\lambda = 4.2$ and 8.3 m.
Mean value $n = 1.4$; standard deviation $= 0.5$.

(ii) *Simultaneous Observations on 72 Mc/s. (4.2 m.) and 46 Mc/s. (6.5 m.)*

A similar series of simultaneous observations on wavelengths of 4.2 m. and 6.5 m. have been made as follows: 12–27 August 1947, 57 observations on the Perseid radiant (11 rejected); 14 December 1947, 79 on the Geminid radiant (17 rejected); 3–4 January 1948, 30 on the Quadrantid radiant (2 rejected).

The distribution in values of the exponent n calculated for each pair from equation (9) is plotted in figure 2 and gives a mean value of 1.7 for n with a standard deviation of 0.6.

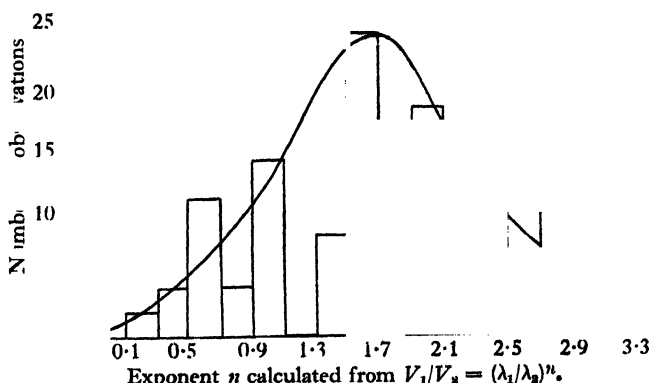


Figure 2. Distribution curve for simultaneous observations on $\lambda = 4.2$ and 6.5 m.
Mean value $n = 1.7$; standard deviation $= 0.6$.

(iii) *Triple Simultaneous Observations on 208 Mc/s. (1.4 m.) 72 Mc/s. (4.2 m.) and 46 Mc/s. (6.5 m.)*

Preliminary attempts have been made to observe the same meteor trail simultaneously on wavelengths of 1.4, 4.2 and 6.5 m. Equations (6) and (7) indicate that the voltage amplitude at the receiver can be expected to decrease as $\lambda^{3/2}$ and hence only the trails with very dense ionization can be expected to give echoes on the 1.4 m. equipment specified in table 1. This expectation has been realized and only four triply observed radio echoes have so far been measured. These yield 12 values for the exponent n , calculated from equation (9) with

mean values as follows : λ 4.2 m. and 6.5 m., mean $n=1.7$; λ 6.5 m. and 1.4 m., mean $n=1.5$; λ 4.2 m. and 1.4 m., mean $n=1.4$.

§ 6. DISCUSSION OF RESULTS

There are two principal uncertainties in these observations. (i) Although geometrically similar aerial systems were used in the comparison equipments, slight differences of terrain and surroundings inevitably introduce variation in the polar diagrams of the aerials. It can therefore be anticipated that a certain number of meteor trails will form in regions where the polar diagrams of the comparison equipments are not exactly similar. (ii) The measurements of signal/noise ratio were made by visual observation of the cathode-ray tube display. Two observers may not agree exactly over the amplitude of a single echo, and for a given observer there will be a certain inconstancy in his estimates of amplitude. Any systematic bias due to the first effect has been counterbalanced by alternating the observers between the equipments but the second effect could not be eliminated in this work.

Both (i) and (ii) will give rise to a considerable scatter of the results about the mean value. In the results described in § 5(a) 40 of the 321 pairs deviated from the mean by more than three times the standard deviation and were rejected. Similarly under § 5(b) 30 of the 166 pairs were rejected.

The final mean values for n are in very reasonable agreement with the predicted value of 1.5 given by equations (6) and (7) over the wavelength range $\lambda=4.2$ to 8.3 m. and the preliminary results on $\lambda=1.4$ m. indicate that the law also holds down to this wavelength.

§ 7. ELECTRON DENSITY IN THE TRAILS

By using equation (7) it is possible to calculate the electron density in the meteor trails, since G, P, λ are constants of the apparatus, and ϵ and R are measured for individual echoes. The measured densities lie in the range 10^9 to 10^{12} electrons/cm. path. The distribution of these densities in the major showers will be dealt with in Part II.

Of particular interest are those cases where visual observations have been made concurrently with the radio echo observations, since it is then possible to measure the electron density of the trails of meteors which are of visual magnitude. Coincidence measurements of this type between a visual meteor and the radio echo on a single wavelength were made during the 1946 Perseid shower (Prentice, Lovell and Banwell 1947), during the 1946 Giacobinid shower (Lovell, Banwell and Clegg 1947) and between a visual meteor and the radio echo on two wavelengths during the 1947 Perseid shower. Detailed associations of visual magnitude with electron density have not yet been made but the results agree in giving electron densities of 10^{10} to 10^{12} electrons/cm. path for the visually observed meteors. The measurements on the 1946 Perseid and 1947 Perseid showers indicate that a 5th magnitude meteor which is on the limit of naked eye visibility produces approximately 2×10^{10} electrons/cm. path and that a very bright meteor (\sim magnitude +1) produces approximately 10^{12} electrons/cm. path. These results are in very good agreement with the theoretical calculations of Herlofson (1948).

§ 8. COMPARISON WITH OTHER THEORIES

Previous calculations of the densities of electron concentrations which give rise to the transient radio echoes have been made by Eckersley (1940) and Pierce (1938).

Eckersley considered that the echoes were due to clouds of electrons and treated the scattering of radio waves as analogous to the scattering of α -particles by heavy atoms. On this theory the voltage amplitude returned to the receiver should be proportional to λ^2 . Eckersley's measurements in the wavelength range $\lambda = 15$ m. to 35 m. showed agreement with this λ^2 law. The measurements described in this paper do not cover this wavelength range but it is evident that the λ^2 law does not apply over the wavelength range $\lambda = 1.4$ to 8.3 m. discussed here. Eckersley's theory does not explain the critical orientation effect of the meteor trails with respect to the aerial beam (Pierce 1938, Hey and Stewart 1946, 1947, Lovell, Banwell and Clegg 1947) and it also yields values for the electron density in visibly observed trails which are in disagreement with Herlofson's calculations (1948). As discussed by Lovell (1948) the transient echo phenomena show a sudden change as the wavelength increases beyond 8 m. Below this wavelength the radio echoes are due to the trails of meteors which are associated with the visible showers originating from definite radiants. In the higher wavelength range the echoes due to these shower meteors are generally submerged in a very high background rate showing a diurnal and seasonal variation (Eckersley 1940, Appleton and Naismith 1947, Eastwood and Mercer 1948). These diurnal and seasonal variations can be explained if the earth is passing through a general distribution of meteoric dust, too minute to produce visible meteors. Eckersley's measurements refer to these latter type of transient echoes, and although the reasons are not yet understood it is possible that the change from the λ^2 to $\lambda^{3/2}$ law may be associated with the change in origin of the transient echoes.

Pierce (1938) calculated the scattered energy from a meteor trail compared with that scattered from an ionized layer, by assuming that the electrons in the trail formed a long column with a diameter *large* compared with the wavelength and with such a high electron density that the radio wave did not penetrate to the inner parts of the trail. These calculations predict the same variation of scattered energy with λ and R as equation (6) and from this point of view the experimental results described here are not in disagreement with Pierce's theory. There are, however, two facts which indicate that Pierce's fundamental assumptions are in error.

(i) Pierce assumes that the meteor trail has already diffused so that its diameter is large compared with λ before the radio echo is observed. Other aspects of the work described in this paper (Part III, see also Ellyett and Davies 1948) show that this assumption is incorrect, at least for observations in the wavelength range $\lambda = 1.4$ to 8.3 m. The echo is obtained from the trail in process of formation as it crosses the beam of the aerial. The subsequent expansion of the electron column by diffusion, to the dimensions assumed by Pierce to be the initial state, can then be shown to be a necessary consequence of the change of duration of the transient echoes with the azimuth of the aerial beam described by Lovell (1948; see also Part III) and explained by Herlofson (1948).

(ii) Pierce's calculations lead to estimates of the electron density in visible meteor trails which differ by orders of magnitude from the densities calculated from equation (6) and are thus in disagreement with contemporary theory

(Herlofson 1948). Also in order to explain the intensity of a meteor echo compared with the intensity of the echo from the F region, Pierce, following Maris (1929) has to assume that the entire kinetic energy of the meteor is spent in ionization, whereas Herlofson (1948) shows that only 10^{-6} of the kinetic energy of the meteor is spent in this manner.

ACKNOWLEDGMENTS

The work described in this paper has been carried out at the Jodrell Bank Experimental Station of the University of Manchester. The authors wish to thank Professor P. M. S. Blackett for his interest and encouragement during the development of the work. The authors are indebted to their colleagues who shared the task of making these measurements, particularly to A. Aspinall, J. G. Davies, C. D. Ellyett, I. A. Gatenby, V. A. Hughes and F. Moran.

REFERENCES

- APPLETON, E. V., and NAISMITH, R., 1947, *Proc. Phys. Soc.*, **59**, 461.
 BLACKETT, P. M. S., and LOVELL, A. C. B., 1941, *Proc. Roy. Soc. A*, **177**, 183.
 EASTWOOD, E., and MERCER, K. A., 1948, *Proc. Phys. Soc.* (in publication).
 ECKERSLEY, T. L., 1940, *J. Instn. Elect. Engrs.*, **86**, 548.
 ELLYETT, C. D., and DAVIES, J. G., 1948, *Nature, Lond.* (in publication).
 HERLOFSON, N., 1948, *Rep. Progr. in Phys.*, **11**.
 HEY, J. S., and STEWART, G. S., 1946, *Nature, Lond.*, **158**, 481; 1947, *Proc. Phys. Soc.*, **59**, 858.
 LOVELL, A. C. B., 1947, *Nature, Lond.*, **160**, 670; 1948, *Rep. Progr. in Phys.*, **11**.
 LOVELL, A. C. B., BANWELL, C. J., and CLFGG, J. A., 1947, *Mon. Not. R. Astr. Soc.*, **107**, 164.
 MARIS, H. B., 1929, *Terr. Magn. Atmos. Elect.*, **34**, 309.
 PIERCE, J. A., 1938, *Proc. Inst. Radio Engrs.*, **26**, 892.
 PRENTICE, J. P. M., LOVELL, A. C. B., and BANWELL, C. J., 1947, *Mon. Not. R. Astr. Soc.*, **107**, 155.

LETTERS TO THE EDITOR

Polarization of Second Order Raman Effect in Alkali Halides

We have succeeded in obtaining Raman spectra of sodium chloride, potassium bromide and potassium chloride which show polarization effects. The spectra obtained for NaCl correspond to the unpolarized spectrum obtained by Rasetti (1931) and by Krishnan (1946). With a Hilger medium quartz spectrograph using 2537 Å. as the exciting line, with a mercury vapour filter to absorb the resonance line after scattering, and a Wollaston prism in the scattered beam, useful spectra were obtained in two days with NaCl and KBr, and in four days with KCl. Single crystals were grown for these experiments.

In order to achieve this, a wide-angled cone of incident radiation was necessary. In practice, no limitation was put on the beam other than that naturally afforded by the critical angle. For high values of depolarization little error results from this, while for low values, for example for KBr, a depolarization of zero would be measured as 0.2. Thus it is clearly possible to distinguish between "polarized" and "depolarized" scattered radiation.

The general appearance of the spectra is that of peaks on a continuous background: this simply means that under the given conditions one could not distinguish between peaks of intensity in a continuum or lines imperfectly resolved. There is a considerable difference between the spectra of NaCl and KBr on the one hand and KCl on the other. The two former were similar in that there were a number of peaks which were partially polarized,

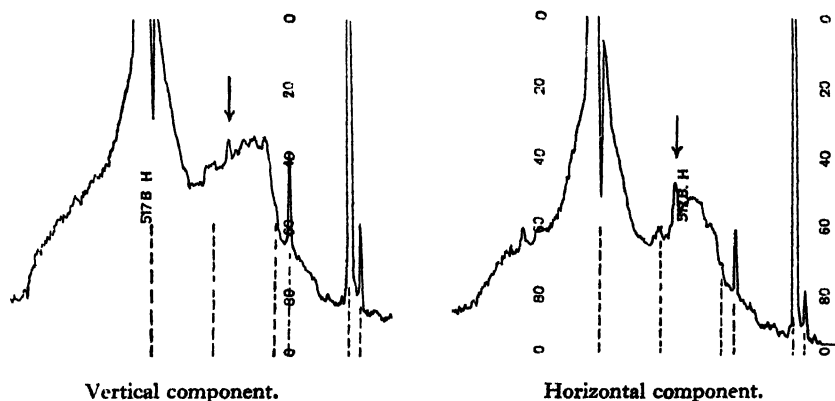
and one outstanding peak depolarized with a depolarization in excess of unity. This peak was at 232 cm^{-1} in NaCl and at 127 cm^{-1} in KBr.

KCl gave a much weaker spectrum, and there was no outstanding depolarized peak; all those strong enough to be observed were well polarized.

It is thought that an explanation can be sought in the approximate equality of mass of the two atoms in KCl and the disparity of mass of the constituent atoms in the other two crystals. On the cyclic lattice theory of Born, it is sufficient to consider the one-dimensional lattice of two atoms to see that peaks of intensity of the continuum present when the ratio of the atom masses differs considerably from unity will fade as the ratio approaches unity. This at once accounts for the relative absence of peaks of intensity in KCl present in NaCl and KBr.

An attempt is now being made to grow a single crystal of RbBr which has a mass-ratio of 1.07 which, if the above explanation is correct, should give polarized spectra similar to those of KCl, which has a mass-ratio of 1.1.

The type of spectrum obtained is seen in the figure, which shows microphotometer curves of the vertically and horizontally polarized spectra for rocksalt. In each case 232 cm^{-1} is



Raman spectrum of rocksalt.

Note: The numbers 517 B.H. in the diagram are merely the recording paper code and are to be ignored.

marked with an arrow. The dip in the high peak is the position of the exciting (resonance) line. The vertical component is privileged in the arrangement used, and this should be remembered in comparing the two curves. The position of mercury lines is indicated by vertical dashed lines.

Adam Hilger Ltd., London N.W. 1.
30 January 1948.

A. C. MENZIES.
J. SKINNER.

KRISHNAN, R. S., 1946, *Proc. Roy. Soc. A*, **187**, 188.

RASSETTI, F., 1931, *Leipziger Vorträge, Molekülstruktur*, p. 59.

REVIEWS OF BOOKS

Theory and Applications of Electricity and Magnetism, by CHARLES A. CULVER.
First Edition. Pp. 594, 407 figs. (London and New York: McGraw Hill Book Co. Inc., 1947.) 25s. net.

This attractive text book was written to provide an introduction to electromagnetism for students at American Universities and "is intended to serve as a basis for advanced study in physics and chemistry and also to lay the foundation for courses in electrical

engineering", but it should also meet the needs of second and third year students of physics and electrical engineering at British Universities.

In its thirty chapters the book surveys the elements of electricity and magnetism, applied electricity, particle physics and electronics. The theoretical approach to electrostatics, magnetostatics and the magnetic effects of currents conforms to the conventional pattern to be found in most text books, but the treatment is clear and comprehensive, with no mathematical demands on the reader beyond a knowledge of elementary algebra and calculus. The electrostatic, electromagnetic and practical systems of units are employed in the text but a welcome passing reference is made to the metre-kilogram-second system of units.

The distinctive feature of the book which calls for special comment is the very large number of practical applications that are described both for their intrinsic importance and to supplement theoretical discussions. To illustrate how the author has justified the word "applications" in the title of the book, it suffices to state that there are excellent chapters on each of the following topics: D.C. measurements (bridges and potentiometers); applications of the thermal effects of currents (electric furnaces, welding, arcs and lighting); electrolysis and primary and secondary cells; thermoelectricity and its uses; the magnetic effect of currents (ammeters and voltmeters).

The book contains one chapter on A.C. theory which expounds the vector method and another short chapter explaining the j -notation with applications. Alternating current instruments claim a chapter and electrical units another.

The last third of the book is a rapid survey of electronics and atomic physics with a stress on practical applications rather than on theoretical details. All this, within the author's terms of reference, is very well done, although the chapter on electricity in gases would seem to indicate that the author has not read sufficiently widely in this subject.

The final two chapters deal respectively with thermionic tubes and their uses and with electromagnetic waves and their applications.

The student's interest is stimulated by frequent reference to the history of ideas and applications but in this the author proves himself by no means infallible. For instance, it comes as a surprise (p. 468) to learn that Moseley worked at the Cavendish Laboratory. On p. 428 we learn, with reference to the nature of cathode rays, that "the German school of thought held, more or less, to the view that the rays consisted of some sort of wave disturbance; while English physicists felt that they were dealing with some type of electrified particles". This statement is inconsistent with the fact that in 1897 both Wiechert and Kaufmann published values of e/m , and that the former was the first to draw the correct conclusion concerning the mass of the cathode corpuscles. On p. 57 it is stated that sun-spots and terrestrial magnetism have been studied in relation to each other for over 1000 years. This is presumably a misprint for 100 years.

These, however, are trivialities. More serious is a diagram (figure 338) of a cyclotron showing the pole pieces spanning scarcely one-third of the diameter of the dees, and a diagram illustrating the principle of radar (figure 399) in which six pulses are shown in flight simultaneously between the ground equipment and the airborne target. It is also time that writers of text books abandoned the practice of attributing the external magnetic field of the earth to an enormous internal bar magnet (figure 24); a doublet source near the centre is a much closer representation. In this context the author remarks: "The terrestrial poles are below the earth's surface". Whether they can be reached by a suitable mining operation is not stated.

Useful sets of test problems accompany most chapters and the book is well bound and excellently printed.

The author ends his book on an optimistic note: "The science of electricity and magnetism has brought countless blessings to mankind. In the years to come it will make the earth a still better place in which to live and have our being". Be this as it may, the author has given practical effect to his belief in writing a book which can give to an imaginative student a vision of what is possible. The book can be confidently recommended.

L. G. H. HUSLEY.

THE PROCEEDINGS OF THE PHYSICAL SOCIETY

VOL. 60, PART 6

1 June 1948

No. 342

A Further Study of the γ -Radiation from Polonium

By B. ZAJĄC, E. BRODA AND N. FEATHER

Department of Natural Philosophy, University of Edinburgh

MS. received 7 February 1948

ABSTRACT. Absorption experiments using lead, gold and tungsten have shown that, in addition to the known γ -radiation of 0.77 Mev. energy, polonium emits other (soft) radiations of which the most intense has a quantum energy of 84 ± 4 kev. The intensity of this radiation is of the same order of magnitude as that of the hard radiation (roughly one quantum per 10^5 disintegrations). Experiments by the recoil method indicate that the emission of the polonium γ -radiation is not delayed by more than 10^{-1} sec. The results of Chang (1946) concerning the fine-structure of the α -particles of polonium remain uncorrelated with all other experimental evidence. Whilst present information regarding γ -ray energies and intensities may be reasonably explained, satisfactory explanation of the α -particle fine structure "data" appears as remote as ever.

§ 1. INTRODUCTION

INTEREST in the γ -radiation of polonium has recently been revived by the publication by Chang (1946) of quite unexpected evidence for a many-lined fine structure in the α -particle spectrum of this body. Two experimental studies of the polonium γ -radiation have since been described (de Benedetti and Kerner 1947, Siegbahn and Slätis 1947) and a critical discussion of Chang's results in relation to earlier information on the γ -radiation (Bothe and Becker 1930, Webster 1932, Bothe 1935, 1936) has been given by one of us (Feather 1946). The experiments now to be described were, like the others already mentioned, prompted by a consideration of Chang's results; they deal more particularly with the question whether a true nuclear radiation of quantum energy less than that of the now (Siegbahn and Slätis 1947) well-established "line" of 0.77 Mev. energy is in fact emitted, and secondly whether or not the emission of the polonium γ -radiation is measurably delayed in relation to the primary α -disintegration process. It should be remarked that the absorption measurements of Siegbahn and Slätis (1947) and de Benedetti and Kerner (1947) would appear to rule out the possibility of any significant emission of quantum radiation in the energy range 0.1 to 0.7 Mev. (or even of somewhat lower energy); on the other hand the only possible remaining way of explaining Chang's results on the basis of nuclear excitation followed by γ -ray emission is to postulate the relatively frequent emission of quanta of low energy (Feather 1946). It may be said in anticipation that no adequate low-energy quantum emission could be established in our experiments, although we have shown, in apparent contradiction of a statement of de Benedetti and Kerner, that weak nuclear radiations of about 84 kev. energy are present. No evidence that any appreciable fraction of the polonium γ -radiation is delayed by as much as 10^{-1} sec. after α -emission was obtained. It will be recalled that the possibility of such delayed emission was discussed in an attempt to interpret (Feather 1946) the reported observation

of γ -radiation of quantum energy 0.2 to 0.43 mev. (Bothe 1935, 1936) in terms of excitation of nuclear states of corresponding energy only to the extent of a few times per million disintegrations. Now that the emission of such γ -radiation has been effectively disproved (de Benedetti and Kerner 1947, Siegbahn and Slätis 1947) the reason behind this suggestion obviously disappears.

§ 2. ABSORPTION EXPERIMENTS

Absorption of the γ -rays in lead was investigated over a range of thickness from zero to 3.4 gm/cm², and in a separate experiment a careful comparison of the absorption in lead, gold and tungsten was made up to thicknesses of 450 mg/cm². The first experiment extended over a period of 15 days, and the second (comparison) experiment over 48 days. No significant discrepancy appeared in the process of correlating all the observations in these experiments by simple correction for decay, taking $\lambda = 5.74 \times 10^{-8} \text{ sec}^{-1}$. The strength of the source at the beginning of the absorption measurements was about 6 mc., and a rough estimate based on β -particle counting showed that at this stage the radium (D + E) contamination was responsible for about 10^{-5} disintegrations per disintegration of polonium. The source was deposited by evaporation of an HCl solution, drop by drop, on a polythene disc 50 mg/cm² thick over an area 0.5 cm. in diameter.

A thin (end) windowed counter was employed for detection of the γ -radiations, the polythene disc carrying the source being supported on wood at a distance of approximately 2 cm. from the counter window, which was 1.4 cm. in diameter and 2.6 mg/cm² thick. The window material was mica. A standard argon-alcohol filling was employed and the counter pulses were registered by a "scale of 32" counting set. The average background counting rate of the (unshielded) counter was about 26 min⁻¹. Throughout the experiments aluminium absorbers of 480 mg/cm² thickness were placed directly over the source to cut out the primary β -particles from the radium E "impurity", and a further absorber of aluminium of 30 mg/cm² was invariably inserted immediately in front of the counter window so that the secondary electrons entering the counter should always originate in the same material. All absorbers were square foils of 3.2 cm. side.

The lead absorption curve obtained with this experimental arrangement is given in figure 1. Analysis of this curve by "least square" and graphical methods indicates component radiations as follows :

- hard component (μ/ρ) = 0.086 cm²/gm., intensity 75%,
- soft component (μ/ρ) = 2.2 cm²/gm., intensity 16%,
- more absorbable radiation, intensity 9%.

The value of the mass absorption coefficient of the hard component is to be compared with 0.0815 cm²/gm. obtained by de Benedetti and Kerner in measurements up to 62 gm/cm², and 0.081 cm²/gm. deduced by Siegbahn and Slätis from observations over 33 gm/cm² absorber thickness *. It should be entirely unnecessary to point out that our value cannot claim anything approaching the independent weight of either of these other two determinations, made with much stronger sources under better geometrical conditions, but the numerical agreement is at least satisfactory. On the other hand, our observation of softer components of quantum radiation, transmitted through 0.5 gm/cm² of aluminium and res-

* As long ago as 1931 rough absorption experiments extending to 51 gm/cm² of lead were made by Blau and Karc-Michailova (1931).

possible for 25% of the total γ -ray effect in our counter, is a matter of more interest. It is explicitly stated by de Benedetti and Kerner that they could detect no radiation softer than the main component of about 0.8 mev. energy until their aluminium absorber thickness was reduced below 0.5 gm/cm². Likewise the lead absorption curve of Siegbahn and Slätis shows little or no trace of any softer component. However, these authors used a counter with a gold cathode, and if this counter was the same as that later employed by Slätis (1947) in another investigation then, from information provided in the later paper, it would appear that absorption of our soft components in the wall of the counter would have amounted to at least 75% in the arrangement used. It cannot be held, therefore,

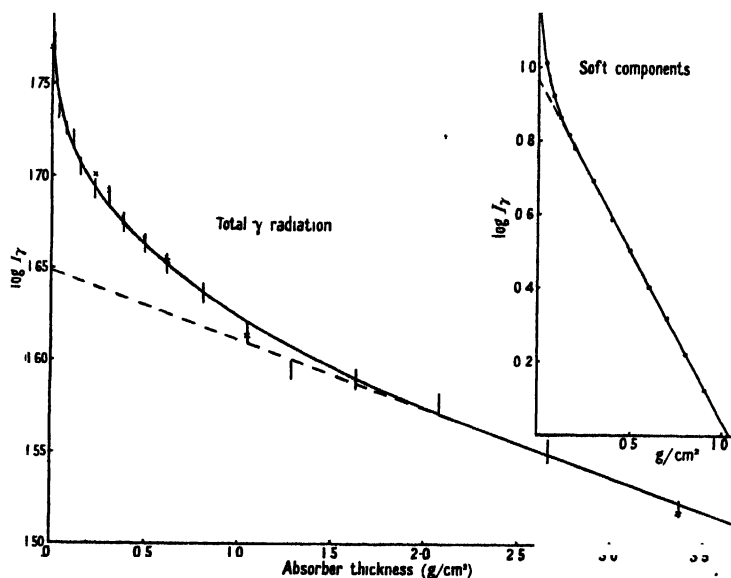


Figure 1

× Observations with 50 mg/cm² polythene above the source.

that serious conflict with the results of Siegbahn and Slätis is established, and without further information as to the degree of attention given to the initial portion of their absorption curve by de Benedetti and Kerner (their published curve shows no plotted point below 1 gm/cm² thickness of lead) it is impossible to say whether there is any outstanding disagreement with their observations, either. It can, however, be stated quite definitely that the total intensity of the soft components observed in our experiments is at least 7 times that of any effect which might be attributed to quantum radiations from the radium (D + E) impurity in the source—and in any case the absorbability of our soft components is quite different from that of the general quantum radiations from a source of radium (D + E + F) in equilibrium. With such a source, under conditions similar to those obtaining in the present experiments, the effective mass absorption coefficient of the radiation transmitted by 0.5 gm/cm² of aluminium is roughly 1.1 cm²/gm. in aluminium, or about 25 cm²/gm. in lead. This is entirely different from $(\mu/\rho)_{Pb} = 2.2$ cm²/gm. which characterizes the main soft component now under discussion.

It should also be stated that there is no possibility that the soft radiations arise from disintegration effects, or as a result of inelastic scattering of the

α -particles, in the aluminium absorber placed over the source. Observations made with 50 mg/cm^2 of polythene directly above the source (see figure 1) showed that these radiations still persisted with the same intensity as before.

The comparison measurements with absorbers of lead, gold and tungsten were carried out in an attempt to reach more definite conclusions concerning the quantum energies of the soft components. On the basis of the lead absorption measurements alone it would be possible to assign an energy of about 145 kev. or 84 kev. to the main soft component, depending on whether or not it were assumed that K -shell absorption was involved. Figure 2 shows the results of the comparison measurements, corrected for decay of the polonium as already described. It is at once evident from the figure that in general the soft radiation is more strongly absorbed in gold and tungsten than in lead—and slightly more so in gold than in tungsten. This result immediately rules out the possibility

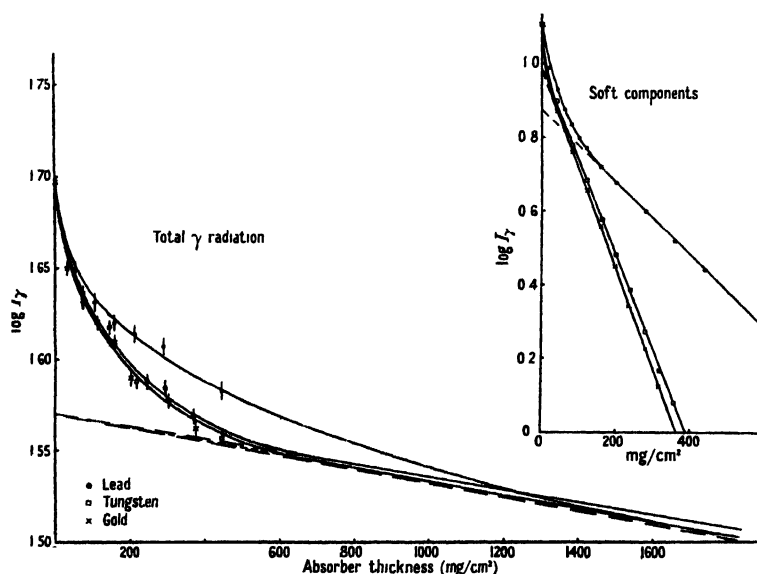


Figure 2.

that the main component of soft radiation has quantum energy greater than 87.9 kev., the K -absorption energy for lead or less than 80.7 kev., the K -absorption energy for gold. A more detailed (graphical) analysis of the curves of figure 2 was carried out on the following assumptions :

- (i) the hard component, deduced from figure 1 as responsible for 75% of the total γ -ray effect in our first experiment, was responsible for the same fraction of the total effect in the comparison experiment (figure 2);
- (ii) the mass absorption coefficients of the hard component in gold and tungsten could be calculated from the value obtained in the first experiment for absorption in lead, using standard theoretical expressions. Values of $0.084 \text{ cm}^2/\text{gm.}$ (gold) and $0.079 \text{ cm}^2/\text{gm.}$ (tungsten) were deduced on this assumption.

The results of this further analysis are summarized in table 1 which includes the corresponding results from the previous analysis of the lead absorption curve of figure 1 for additional comparison.

Table 1

Experiment	I	II	II	II
Absorber	Pb	Pb	Au	W
Main soft component	$\left\{ \begin{array}{l} \mu/\rho \text{ (cm}^2/\text{gm.)} \\ \text{effective intensity (\%)} \end{array} \right.$					2.2	2.2	6.1	5.7
More absorbable radiation—effective intensity (%)	16	15	19	19
	9	10	6	6

It is at once evident from the figures given that it is impossible to provide a complete explanation of all these results in terms of two monochromatic soft components only; the term "main soft component" which has been used hitherto has obviously no unique reference to a single radiation. This latter conclusion derives from two considerations, first that the apparent intensity of this hypothetical component is greater on the basis of gold (or tungsten) absorption than on the basis of lead absorption measurements, and secondly that corresponding to $(\mu/\rho)_{\text{Pb}} = 2.2 \text{ cm}^2/\text{gm.}$ we should expect $(\mu/\rho)_{\text{Au}} = 8.4$ and $(\mu/\rho)_{\text{W}} = 6.8 \text{ cm}^2/\text{gm.}$, instead of 6.1 and 5.7 $\text{cm}^2/\text{gm.}$ respectively. Clearly there must be present radiation more absorbable than the most prominent soft component radiation according to lead absorption measurements, but less absorbable than that component in gold and tungsten. Such a radiation must obviously have a quantum energy greater than 87.9 kev., the K -absorption energy of lead. On the other hand, the presence of a radiation more absorbable in all three elements than the most prominent soft component must also be postulated. A quantum energy of less than 35 kev. would thus appear to be indicated. We may therefore accept a very rough three-component interpretation of the soft radiation as follows (table 2)

Table 2

Quantum energy (kev.)	35	84 \pm 4	105 \pm 15
Effective intensity (%)	6	15	4

Further discussion of this interpretation is reserved for a later stage, here it should merely be re-emphasized that the figures given under heading "effective intensity" in table 2 represent percentages of the total γ -ray counting rate in our experimental arrangement, not absolute, or even relative, values of true quantum intensities per disintegration of polonium.

§ 3. DELAY EXPERIMENTS

These experiments were designed in the first instance to investigate the possibility that the residual nucleus ^{206}Pb , which results from the α -disintegration of polonium, is occasionally produced in a metastable state of lifetime of the order of minutes; they were later extended to search for possible γ -ray lifetimes as short as a few milliseconds. The first experiment originated in the suspicion (Feather 1946) that the unidentified activity of 1.6 m. half-value period, which Waldman and Collins (1940) have reported as produced by the action of high-energy x-rays on lead, might result naturally in a rare mode of the disintegration $^{210}\text{Po} \rightarrow ^{206}\text{Pb}$. As already stated, this experiment gave a negative result. It was carried out by exposing a brass disc collector of 1.4 cm. diameter at a distance of 1 cm. and at a potential of -700 v. with respect to a "clean" source of about 12 mc. of polonium deposited on a platinum foil of 4 cm^2 area. Exposures lasted for a standard time of 12 minutes. Any activity collected by the disc was then observed over a period of 10 minutes immediately following the end of the exposure, using a thin-windowed counter and an absorber of 15 mg/cm^2 paper to cut out the effect of the α -radiation of the polonium which was collected

by aggregate recoil. As a result of 14 exposures it was concluded that the collector carried no activity giving an initial counting rate of more than 1 count per minute under our experimental conditions. The α -activity of the polonium collected by aggregate recoil was estimated, by removing the paper absorber, as 1.4 min^{-1} per minute exposure in the same geometrical arrangement, this figure corresponding roughly to one polonium atom collected per 3000 α -disintegrations of the source.

The next experiment, on possible short-period delays, was carried out in two different arrangements exemplifying the same basic principle. This principle was to examine the activity of a collector permanently exposed to a polonium source, the source and collector being so disposed in the neighbourhood of the end-windowed counter that the counter was considerably more sensitive to γ -radiation emitted by atoms disintegrating on the collector than to γ -radiation from the primary source.* Transference of recoiling atoms from the source to the collector in the applied field was calculated to take place in an average time of $2.4 \times 10^{-3} \text{ sec.}$ in the first arrangement and $5 \times 10^{-2} \text{ sec.}$ in the second. In spite of the longer transference-time the second arrangement had the great advantage over the first that the collector subtended a much greater solid angle at the counter window than before: in the second arrangement, in fact, the collector was separated from the window only by a 0.5 cm. air gap and an aluminium absorber 30 mg/cm^2 thick. In each arrangement the counting rate with source and collector in position was determined both when the source and collector were at the same potential and also when the collecting field was applied. A "genuine" increase in the counting rate on the application of the field would in this experiment provide clear evidence for the production of an activity in the recoil nucleus ^{206}Pb of any lifetime between a few milliseconds and 1 minute. In fact, the mean increase in three sets of observations using the first arrangement, was $0.80 \pm 0.53 \text{ min}^{-1}$ and, in five sets of observations using the second arrangement, $0.40 \pm 0.56 \text{ min}^{-1}$. The γ -ray counting rate determined with the primary source in the collector position in arrangement (i) was $20.5 \pm 2.4 \text{ min}^{-1}$, and it was estimated that in arrangement (ii) it would have been at least 60 min^{-1} . (As already implied a direct determination of the γ -ray effect of the polonium source through only 30 mg/cm^2 aluminium could not easily be made on account of the radium (D + E) impurity). The efficiency of collection of ^{206}Pb recoil atoms was estimated—in the only way available, in terms of the efficiency of aggregate recoil collection—as about $\frac{2}{3}$ of the efficiency in the standard exposure arrangement of the long-period delay experiment (see above) in arrangement (i), and as about $\frac{1}{3}$ of this standard efficiency in arrangement (ii). The latter result was all the more gratifying since the average distance travelled by a recoil atom in the field—and that not in a straight line—was about 10 cm. in this arrangement, as against 1 cm. under "standard" conditions.

§ 4. DISCUSSION

In discussing the experimental results described above we may deal first, quite briefly, with the results of the delay experiments. Always assuming that the recoil efficiency of the primary polonium source was high—which the appearance of the activated platinum foil and our observations on aggregate recoil might be held to justify—it is obvious that these results show that not more

* Essentially the same principle was used by Ellis (1932) and also by Wright (1932) in work on ThC and RaC', respectively.

than 10% of the γ -ray effect in our counter arises from activities of lifetime greater than 5×10^{-3} sec. and not more than about 3% of the total effect from activities of lifetime greater than 10^{-1} sec. It is significant to remark that this second statement includes γ -ray activities which because of a high degree of internal conversion might show up only through the emission of L x-rays (Curie and Joliot 1931), since the aluminium absorber in the second arrangement was sufficiently thin to transmit a considerable fraction of such fluorescent radiation.

Turning to the results of the absorption experiments one important conclusion can be stated at the outset. Because of the close similarity of the absorption curves obtained with gold and tungsten absorbers it must be concluded that no significant component of the "soft" radiation from polonium has quantum energy intermediate between the K -absorption energies of tungsten and gold, i.e. in the energy range between 69.4 and 80.7 kev. Now the $K\alpha_1$ and $K\alpha_2$ radiations of elements 81 to 84 inclusive lie within this range, thus we have the important result that the soft radiations which we have detected do not include the K -radiations of lead. That these radiations must be emitted following the internal conversion of the 0.77 Mev. polonium γ -ray is undeniable, but they are evidently below the limit of detection with our arrangement: present considerations, in fact, support previous estimates of the absolute intensity of the photoelectron lines (Feather 1946) and show that the internal conversion coefficient for the 0.77 Mev. radiation cannot be abnormally large (not greater than about 10^{-2}). Since they are not K x-rays obviously the soft radiations of table 2 must be predominantly of nuclear origin. Before we can make much progress towards an understanding of their mode of emission, however, we must make some attempt to estimate their intensity. This can only be done in the present instance by a rough comparison with the harder radiation.

Let us consider first the intensity of the main hard component. The "best" estimate in this case would appear to be $(1.0 \pm 0.4) \times 10^{-5}$ quanta per disintegration (Feather 1946)—and no difficulty would ensue if this were the only γ -radiation emitted. For, if the same change of spin were involved in an α -disintegration leading to the 0.77 Mev. excited state as in that leading to the ground state of the nucleus ^{206}Pb , we should expect an excitation of the 0.77 Mev. state of $(1.8 \pm 0.5) \times 10^{-5}$ per disintegration (Feather 1948). The agreement between this figure and the observed γ -ray intensity is remarkably close, and in the absence of the soft radiation we might well be satisfied that the rare nuclear processes were completely understood. We have now to see whether it is possible to explain at the same time the emission of the softer γ -radiations in terms of the excitation of other states having spin values consistent with our observation that the soft γ -radiation is not appreciably "delayed". In this connection we should note the estimate of Curie and Joliot (1931) that the intensities of the L - and M -radiations of polonium (or lead) which they observed were about 4×10^{-4} and 1.5×10^{-3} quanta per disintegration, respectively.

Now it is certain from our results that the soft radiations are not considerably more intense (in quanta emitted per disintegration) than the hard component. Using the results of Bradt and collaborators (1946) in making our comparison we should in fact conclude that our 84 ± 4 kev. soft component has roughly the same intensity as the 0.77 Mev. (hard) radiation.* Our other soft components

* Bradt *et al.* found that an aluminium counter is roughly five times as sensitive to γ -rays of 0.8 Mev. energy as to 0.1 Mev. radiation.

would then appear as less intense than this. As concerns the main soft component two possibilities must now be considered: either the major portion of the *L*-radiations of Curie and Joliot follows the internal conversion of this soft γ -radiation (of 84 ± 4 kev. energy) or it does not. On the first assumption the 84 kev. nuclear transition has an intensity of roughly 4×10^{-4} per disintegration, the radiation has an internal conversion coefficient of about 40, and because of intensity considerations it cannot be coupled in any way with the emission of the 0.77 mev. radiation. Separate excitation of a nuclear state of 84 kev. energy has thus to be assumed, and a difference of spin between excited and ground states of some 5 units is necessary to explain the intensity. The assumption has therefore to be ruled out if the radiation is not appreciably delayed.

The second assumption avoids this particular difficulty by allowing coupling of the hard and soft radiations—and so permitting a single high energy state excitation. We might, in fact, on this view, imagine that the 0.77 mev. state was fed chiefly by γ -ray transitions from a state of 84 kev. higher energy (and possibly to a smaller extent from other states in the same energy range) excited directly by α -disintegration. The difference of 84 kev. in the energy available for α -disintegration would be expected to introduce an adverse factor of no more than 3.5 in the partial disintegration probability (which would not seriously disturb the numerical agreement already noted), and if it were assumed, for example, that the ground state of ^{206}Pb and the excited state of 0.85 kev. each had zero spin, and the 0.77 mev. state had spin of 2 units, no major interpretative difficulty need remain—except the explanation of the intensity of the *L*- and *M*-radiations observed by Curie and Joliot and the whole problem of the fine-structure spectrum of Chang (1946). Concerning the first difficulty it might be possible to accept the suggestion, frequently put forward in the past, that the fluorescent radiations arise in the (infrequent) internal ionization of the radioactive atom by the escaping α -particle, having no connection, therefore, either with the appearance of discrete lines in the α -particle energy spectrum, or with nuclear γ -ray transitions, but the problem posed by Chang's results would seem as far from solution as ever. Amongst the rare effects associated with the polonium disintegration only the fluorescent x-ray emission itself appears to have an intensity of the same order of magnitude as the fine-structure lines in Chang's spectrum.

REFERENCES

- DI BENEDETTI, S., and KERNER, E. H., 1947, *Phys. Rev.*, **71**, 122.
 BLAU, M., and KARA-MICHAILOVA, E., 1931, *Sitzungsber. Akad. Wiss. Wien*, II a, **140**, 615.
 BOTHE, W., 1935, *Z. Phys.*, **96**, 607; 1936, *Ibid.*, **100**, 273.
 BOTHE, W., and BECKER, H., 1930, *Z. Phys.*, **66**, 307.
 BRADT, H., GUGELOT, P. C., HUBER, O., MEDICUS, H., PREISWERK, P., and SCHFRER, P., 1946, *Helv. Phys. Acta*, **19**, 77.
 CHANG, W. Y., 1946, *Phys. Rev.*, **69**, 60.
 CURIE, I., and JOLIOT, F., 1931, *J. Phys. et Radium*, **2**, 20.
 ELLIS, C. D., 1932, *Proc. Roy. Soc. A*, **136**, 396.
 FEATHER, N., 1946, *Phys. Rev.*, **70**, 88; 1948, *Nuclear Stability Rules* (Cambridge: University Press) (in preparation).
 SIEGBAHN, K., and SLÄTIS, H., 1947, *Arkiv Mat. Ast. Fys.*, **34 A**, (15).
 SLÄTIS, H., 1947, *Arkiv. Mat. Ast. Fys.*, **35 A** (3).
 WALDMAN, B., and COLLINS, G. B., 1940, *Phys. Rev.*, **57**, 338.
 WEBSTER, H. C., 1932, *Proc. Roy. Soc. A*, **136**, 428.
 WRIGHT, P., 1932, *Proc. Camb. Phil. Soc.*, **28**, 128.

The Semi-Diurnal Variation in Cosmic Ray Intensity

BY PHYLLIS NICOLSON AND V. SARABHAI

Cavendish Laboratory, Cambridge

MS. received 2 September 1947

ABSTRACT. The experimental data on the daily variation in cosmic-ray intensity are first surveyed. It appears to be established by Rau's experiments that the energetic meson component at sea level shows a marked semi-diurnal variation which is in phase with the semi-diurnal variation shown by the barometric pressure. Several authors have suggested that this phenomenon is explicable in terms of the Pekeris theory of atmospheric oscillations. The implications of this explanation on the process of meson formation are here examined quantitatively. It is concluded that the explanation is only possible if mesons arise mainly at about 60 or 70 km. above sea level, which is highly unlikely since the corresponding cross-section for meson production would be much larger than appears plausible.

In addition the experimental data on the diurnal variation are briefly discussed. This variation seems to be more complicated in character than is usually assumed and not attributable to factors such as a heliomagnetic field and fluctuations in the geomagnetic field, which are often held responsible.

§ 1. INTRODUCTION

IF the experimentally observed solar daily variation in cosmic-ray intensity is analysed into Fourier components, it is usually found that the only important components are the diurnal (24-hour period) and the semi-diurnal (12-hour period). It is with the semi-diurnal component that we are chiefly concerned here.

Certain observers (Regener and Rau 1939, Ehmert 1939, Rau 1940, Sarabhai 1945) have found a daily variation in meson intensity which is mainly semi-diurnal and shows a strong positive correlation with that in atmospheric pressure, and have pointed out that the theory of atmospheric oscillations outlined by Pekeris (1937) and described in § 3 below provides at least a qualitative explanation of this phenomenon, which is difficult to understand on any other basis. If this explanation is valid, then the experimental data on the amplitude of the cosmic-ray effect, together with the quantitative predictions of the Pekeris theory, provide a means of determining the approximate height at which mesons are mainly produced and hence the approximate value of the average cross-section for meson production. At the present stage of theory and experimental data relating to meson creation, even a very rough estimate of the value of this cross-section would be useful.

It was clear from the general features of the oscillation predicted by Pekeris that this method of explaining the semi-diurnal variation in meson intensity would certainly lead to rather a large value for the cross-section, but whether this value would be too large to be plausible could not be decided without quantitative investigation. The main purpose of this paper is to describe such investigations and their result (§ 3). One of the subsidiary purposes is to make clear the implications of the Pekeris theory, for the results of this theory have been incorrectly interpreted by certain writers on the semi-diurnal variation in cosmic-ray intensity.

Since there is much apparently conflicting evidence on the daily variation of cosmic radiation, it is essential to survey all the experimental data before proceeding to consider theories to explain any part of these data. We shall now make such a survey and then consider what facts may be looked upon as experimentally established.

§ 2. EXPERIMENTAL DATA

It is important to classify the various experimental results according to the nature of the intensity to which they refer, for the daily variation may be expected to depend upon the constitution of the radiation observed. For example the mesonic and electronic components may not be affected in the same way by short-period fluctuations in meteorological conditions, since these fluctuations may be entirely different in character in the upper atmosphere where the mesons mainly arise and in the lower atmosphere where most of the electrons recorded at sea-level arise.

In addition it is important to make clear whether the observer in presenting his results has tried to allow for changes in meteorological factors such as temperature and pressure, for it often happens that the so-called "corrections" applied are so large that they completely change the character of the daily variation. For example, the "external temperature" correction, applied by many workers in an attempt to eliminate the effect of fluctuations in the temperature of the air above the recording apparatus, may be sufficient to shift the phase of the diurnal variation by eight hours; it is therefore important to consider the justification for this correction. The correction is usually made before harmonic analysis is carried out, by adjusting the hourly recordings by means of a temperature coefficient determined from the correlation of long period (usually average monthly) variations of intensity with temperature. Now the variations in monthly mean intensity at sea level are probably caused mainly by changes in the average height of origin of mesons; this height depends upon the temperature distribution throughout the atmosphere below it and also on the pressure at sea level. Consequently, even in these long period variations, there is not always a simple dependence of the intensity on the temperature at the ground; even if there were it would have little bearing on the relation between hour-to-hour changes in intensity and ground temperature. Test made by Forbush at Cheltenham (U.S.A.) suggest that at this location at least, no external temperature correction ought to be made. Consequently the value of results to which this correction has been applied seems very doubtful.

The practice, often adopted, of correcting hourly readings for fluctuations in barometric pressure by using a barometric coefficient based on comparison of the mean daily intensity and barometric pressure, may not be entirely justified either; for although the barometric effect is certainly due partly to mass absorption, it may also depend upon factors such as movement of meson-producing layers, which are differently related to short and long period pressure changes. A third correction is sometimes applied for fluctuations in the internal temperature of the apparatus, but in the best designed experiments this is made unnecessary by thermostatic control.

2.1 Collection of results

Results are collected in table 1. The notation used therein is as follows :—

λ = geomagnetic latitude of the station.

h = altitude of station in metres above sea level.

ϕ = angle of cone formed by observed radiation; since information on this is not very precise it will be classified simply as wide or narrow.

θ = inclination of axis of observed cone to the vertical.

$D(b, c)$; $S(b, c)$ = diurnal (semi-diurnal) variation of amplitude $b\%$, and maximum at c hours. In cases where the harmonic dial method of analysis has been used the probable error in b is also given.

+ S ; - S imply that the maximum of the variation is displaced by less than one hour from the maximum (minimum) of the semi-diurnal pressure variation.

I.C.; G.C. indicate use of ionization chamber; Geiger counter apparatus.

B.; E.T.; I.T. indicate correction for barometer effect; external temperature effect; internal temperature effect.

2.2 Discussion of the experimental evidence on the S variation

In view of the smallness of the daily variation, observations over a very long period are required before one can be certain that the variation obtained is of physical significance. For this reason the reliability of many of the results quoted above is somewhat doubtful. In the case of the S variation the only results of convincing statistical significance are those of Duperier (1945), based on three years' observations and those of Rau (1939), although the results found by Hoerlin under 10 cm. Fe do carry some conviction on account of the large amplitude of the S effect recorded and the closeness with which it follows the pressure variation.

Rau used two independent ionization chambers of different size suspended under 40 metres of water in a narrow vertical fissure in the rock forming the floor of Lake Constance. The dimensions of this fissure were such that the intensity observed was mainly vertical. The data from the two chambers showed a similar variation. The results given above were obtained with the larger chamber. The first result $S(0.18 \pm 0.04, 11^h 30)$ was found by a harmonic dial analysis of the data recorded on 192 days, while the second result $S(0.14, 10^h)$ was obtained by taking the average of the hourly readings over a period of 15 months, and then harmonically analysing the resulting daily variation curve. The two results when considered together, suggest that the amplitude lies between 0.14 and 0.18%, and we shall in numerical work take 0.16% as the amplitude established by Rau.

The interpretation of Rau's results is much less ambiguous than that of Duperier's for the following reasons: (i) Only mesons are measured instead of the total intensity made up of mesons and soft component. (ii) The angle ϕ is small and consequently the radiation can be regarded as entirely vertical to a good approximation. (iii) Only particles of initial energy $> 10^{10}$ ev. can penetrate the absorber above Rau's apparatus (50 m. water including the atmosphere), which means that the effect of magnetic variations has been eliminated. This is confirmed by the absence of any change in intensity during magnetically disturbed days in the case of Rau's experiments.

In view of the difference in nature of the intensity measured in the two cases, it is not surprising that the variations found by Rau and Duperier are of very different character.

At the equator there are no completely convincing results. Thompson's are uncertain on account of their short duration (only about 6 day's readings for a zone $2\frac{1}{2}^\circ$ in latitude) and the fact that a relatively large correction is made necessary

Table 1

(i) Variation of radiation

Observer	λ	h	Apparatus	ϕ	θ
Barnothy and Forro (1939)	47° N.	124	G.C.	wide	0
Duperier (1945)	54 N.	0	G.C.	wide	0
Alfvén and Malmfors (1943)	58 N.	0	G.C.	narrow	30° N. 30° S. 60° N. 60° S.
Hoerlin reported by Rau (1940)	2 N.	5500 to 6100	I.C.	wide	0
		2500 to 5500	I.C.	wide	0
Kolhorster (1941) (1)	52.4 N.	0	G.C.		45° N. 45° S. 45° E. 45° W.
Sarabhai (1945)	3 N.	900	G.C.	wide	0

(ii) Variation of total

Steinmaurer (1935)	47 N.	590	I.C.	wide	0
Hess and Graziadei (1936)	47 N.	2300	I.C.	wide	0
Doan (1936) (2)	52 N.	150	I.C.		0
Forbush (1937) (3)	50 N.	72	I.C.	wide	0
Schonland, Delatizky and Gaskell (1937) (4)	32 S.	122	I.C.	wide	0
Thompson (1938) (5)	40-55 N. 25-40 N. 10-25 N. 10° S.-10° N. 10-25 S. 25-40 S.	0	I.C.	wide	0
Sarabhai (1945)	3 N.	900	G.C.	narrow	0

(iii) Variation of

Barnothy and Forro (1936) (6)	47° N.	124	G.C.		0 50°
Rau (1939) (7)	50 N.	360	I.C.	narrow	0

(iv) Variation of

Barnothy and Forro (1939)	47° N.	124	G.C.		0
---------------------------	--------	-----	------	--	---

- (1) Results not Fourier-analysed. Appreciable variation with θ ; time of maximum varies from about 7 to 15h.
- (2) Very small barometric variation on 10 days selected. Results from 7 identical recording meters running simultaneously.
- (3) D variation well established.
- (4) Results made uncertain by large temperature fluctuations of apparatus itself.
- (5) Recording made during 12 voyages. On average 6 days readings for each zone $2\frac{1}{2}^\circ$ in latitude. Uncertainty introduced by necessity of correcting for latitude effect.

under little or no absorber

Absorber	Duration of Expts.	Corrections applied	Results	
			D	S
0	2 months	B.		
0	3 years	None B.	(0.25, 17 ⁿ 20) (0.22, 17 ⁿ 20)	-(0.18, 03 ^h 12)
0	2½ years	B.	(0.12 ± 0.03, 10 ^h) (0.08 ± 0.03, 13 ^h) (0.16 ± 0.02, 13 ^h) (0.17 ± 0.02, 19 ^h)	
0	10 days	B., I.T.		-(0.06, 04 ^h 30)
10 cm. Fe	8 weeks	B., I.T.		+ (0.78, 10 ⁿ 47)
0	8 months	None		
0	15 days	None	(0.3, 14 ^h 30)	

meson component

10 cm. Pb	9 months	B. B., E.T.	(0.25, 8 ^h) (0.25, 12 ^h)	
12 cm. Pb	3 years	B., E.T.	(0.2, 10 ^h 40)	(0.5, 00 ⁿ)
12 cm. Pb	10 days	B.	(0.19 ± 0.04, 09 ^h)	
12 cm. Pb	273 days	B.	(0.17 ± 0.02, 11 ^h)	(0.04 ± 0.01, 00 ^h 30)
12 cm. Pb	3 years	B. B., I.T. B., I.T., E.T.	(0.2, 16 ^h) (0.2, 6 ^h) (0.2, 14 ^h)	
12 cm. Pb	11 months in all	B.	(0.33, 14 ^h 20) (0.22, 14 ^h 30) (0.26, 14 ^h 20) (0.23, 13 ^h 50) (0.21, 14 ^h 30) (0.17, 13 ^h 10)	
8 cm. Pb*	12 days	None	0	(3.0, 10 ⁿ 30)

energetic meson intensity

36 cm. Pb	4 months	B.		
36 cm. Pb	3 months	B.		
40 m. H ₂ O	15 months	None	0	(a) + (0.18 ± 0.04, 11 ^h 30) (b) + (0.14, 10 ⁿ)

electronic shower intensity

?	12 months	B.	(0.41, 15 ^h)	
---	-----------	----	--------------------------	--

(6) Curves suggest reversal of phase of S variation as θ changes from 0 to 50, but physical significance of these results is not convincing.

(7) (a) Harmonic dial method used on 192 days observations.

(b) By harmonic analysis of mean of 15 months observations. A second smaller, independent I.C. was used simultaneously and showed similar variation.

* In this case the method of shower anticoincidences was actually used for excluding the soft component, but the arrangement is essentially equivalent to a counter telescope with about 8 cm. Pb absorber.

by the latitude change in a 24-hour period. Sarabhai's experiments are unfortunately also too brief to give convincing results; they suggest the presence of a large +S variation. Hoerlin's results, obtained during the 1932 Cordillera Blanca Expedition and re-examined by Rau (1940), carry the greatest conviction. They suggest very strongly that the variation under an absorber of 10 cm. Fe is semi-diurnal, with large amplitude and phase strikingly close to that of the semi-diurnal pressure variation (meson maximum at $10^{\text{h}}47$, pressure maximum at $10^{\text{h}}50$). It seems likely, therefore, that at the equator the meson component shows a large +S variation but further experiments are necessary to establish this.

However this may be, Rau's conclusion that at latitude 50° the energetic mesons are characterized by a +S variation, appears to have good foundation and it seems worth while to examine quantitatively whether such a variation can be explained in terms of the Pekeris theory of atmospheric oscillations, particularly as this explanation of the effect has been given in several independent papers.

§ 3. EXPLANATION OF THE +S VARIATION

A predominant semi-diurnal variation is not likely to result either from temperature or from magnetic variations, as both would produce mainly diurnal rather than semi-diurnal effects. Apart from this consideration, Rau's results show clearly that magnetic fluctuations are not responsible, as the particles he observes have energies too high to be influenced by them*.

It has already been pointed out independently by Rau and Ehmert (1940) and by Sarabhai (1945) that a +S variation would result from atmospheric oscillations of the kind predicted by Pekeris, provided that the bulk of the mesons arise above 30 km., at which height the Pekeris oscillation has a nodal surface. Before working out the implications of this interpretation we will briefly indicate the nature of these atmospheric oscillations.

3.1 *The theory of atmospheric oscillations*

The diurnal component of the daily variation in atmospheric temperature is considerably larger than the semi-diurnal component, as one would expect. In spite of this, however, in the variation of the barometric pressure at sea-level, the semi-diurnal component is dominant, being about $1\frac{1}{2}$ times as large as the diurnal component. An explanation of the large semi-diurnal pressure variation has long been sought. Kelvin suggested that the atmosphere has a free oscillation of period about 12 hours, so that the semi-diurnal tide excited by the sun is magnified by resonance. Since then many attempts have been made to determine the period of free oscillation of the atmosphere.

G. I. Taylor (1929, 1932, 1936) showed that this period can be calculated from the speed of propagation of waves of explosion such as that caused by the Krakatau eruption in 1883, and that for this eruption the speed with which the wave was propagated corresponds to a free period of about $10\frac{1}{2}$ hours; a similar speed was found by Whipple (1934) for the wave of the Great Siberian Meteor. A period of $10\frac{1}{2}$ hours is, however, fatal to the resonance theory of atmospheric oscillation. In 1937 Pekeris extended the theory which had been outlined by Taylor in 1936,

* In the case of Rau's experiments, tides in Lake Constance provide a possible cause for an S effect, since any change in surface level of the lake alters the thickness of the water absorber above the apparatus. The change in intensity produced by such tides appears to be negligible however, as their magnitude is very small indeed—the change in level from high to low water being of the order of 0.1 inch.

and was able to prove that, when the temperature distribution with height suggested by experimental data is assumed, the atmosphere can have a free oscillation of 12-hour period in addition to one of $10\frac{1}{2}$ -hour period.

As early as 1924 Chapman discussed the semi-diurnal oscillation excited partly tidally and partly thermally by the sun, and concluded that if the atmosphere has a natural period of 12 hours, then this oscillation will be magnified about a hundred-fold by resonance. Recently a more detailed investigation of the magnitude of the semi-diurnal pressure oscillation at different heights has been carried out by K. Weekes and M. V. Wilkes, who have extended the basic theory outlined by Pekeris and have used the Cambridge Differential Analyser to obtain numerical solutions of the equations involved. An account of this work together with detailed numerical results is to be published shortly. Meanwhile, Weekes and Wilkes have very kindly made their results available to the present writers.

The mode of oscillation of 12-hour period is characterized by a horizontal nodal surface at a height of about 30 km., the atmosphere above and below this level swinging in opposite directions. The changes in pressure and the horizontal velocity of winds caused by the oscillations are reversed in phase in crossing this level; the vertical velocity of winds has the same phase at all levels.* The amplitude of the pressure oscillation is small and almost uniform up to nearly 30 km.; at high altitudes it increases rapidly with increasing rarefaction of the atmosphere. The feature of particular significance in considering the effect of such pressure oscillations upon meson intensity at sea level is the variation Δh in the height of an isobaric surface situated at mean height h . Table 3 shows the values of Δh as a function of h calculated by Weekes and Wilkes for an atmosphere having the temperature distribution given in table 2, a distribution based as far as possible upon experimental data. The values given for Δh correspond to an oscillation of amplitude 1 mm. Hg at sea level at the equator, which is very close to that experimentally observed. The amplitude of the oscillation decreases with increasing latitude by a factor approximately $\cos^3 \lambda$; Δh is given for $\lambda = 0$ and 50° .

Table 2. The variation of temperature with height adopted in the calculation of table 3

Height (km.)	Temperature ($^\circ$ K.)	Evidence
0-10	288-220 ; linear fall.	} Up to about 25 km. well established by radio-sonde data. Inferred from the abnormal propagation of sound (Whipple 1935).
10-30	220	
30-55	220-350 ; linear rise.	
55-60	350	} Suggested by radio exploration (Budden, Ratcliffe and Wilkes 1939)
60-77	350-190 ; linear fall.	
77-100	190	

Table 3. The semi-diurnal oscillation of isobaric levels at various altitudes

h (km.)	0	10	20	30	40	50
Δh (km.)	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} +0.010 \\ +0.003 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} +0.012 \\ +0.003 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} +0.012 \\ +0.003 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -0.002 \\ -0.001 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -0.086 \\ -0.023 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -0.272 \\ -0.072 \end{array}$
h (km.)	60	70	80	90	100	
Δh (km.)	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -0.49 \\ -0.129 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -0.51 \\ -0.135 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -0.60 \\ -0.160 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -1.05 \\ -0.28 \end{array}$	$\left\{ \begin{array}{l} \lambda=0 \\ \lambda=50^\circ \end{array} \right. \begin{array}{l} -2.22 \\ -0.59 \end{array}$	

* It is interesting to note that this type of oscillation is compatible with that suggested by the "dynamo" theory developed by Chapman (1919) to account for the semi-diurnal variation of the earth's magnetic field. Chapman finds that the pressure oscillations in the upper conducting layer where the dynamo effect is produced are nearly 180° out of phase with those observed at the ground.

These values do not agree with those calculated by Mailvaganam (1946). From the following considerations it seems that Mailvaganam's results cannot be valid. It can be seen from the results of Pekeris' own analysis that near resonance the horizontal velocity is proportional to the pressure oscillation at any height, and consequently that the curve given by Pekeris (1937, figure 1) for the horizontal velocity also represents the pressure oscillation. Now the pressure oscillation calculated by Mailvaganam does not agree even in character with that given by Pekeris. The most obvious discrepancy is that Mailvaganam's results do not show the node at 30 km. which characterizes all Pekeris' results. One of the causes of the discrepancy seems to be that Mailvaganam has used incorrect numerical values for constants (C and D in his notation). We are indebted to Weekes and Wilkes for investigations on this point.

It is apparent from the nature of the oscillation described above that mesons created above 30 km. will show a + S variation at sea level, for if the isobaric levels in the lower atmosphere move down, corresponding to a fall in sea level barometric pressure, isobaric levels above 30 km. move up; mesons are therefore formed at higher levels and the meson intensity at sea level is decreased as the probability of decay is enhanced. Also the + S variation arising in this way should be much more pronounced at the equator than at latitude 50° , as the pressure oscillation increases by about a factor 4 in going from $\lambda = 50^\circ$ to $\lambda = 0^\circ$. The experimental data at present available at the equator are compatible with the occurrence of such an increase in amplitude of the S variation, although not sufficient to establish it.

We shall now calculate the average height at which mesons would have to arise to produce a + S variation of the magnitude observed by Rau.

3.2 Calculation of the average height of creation of energetic mesons

3.2.1 Treatment assuming all mesons arise at one height

In this calculation we shall assume that all mesons arise at height X above sea-level, and that at this height they have a differential energy spectrum proportional to E^{-n} . It seems probable that mesons at a high level will have a spectrum of this form; the assumption is unlikely to introduce a large error, as the results do not appear to depend critically upon the energy spectrum. Usually n will be taken equal to 3, but one case with $n = 2$ has been worked out to test the dependence of the result upon n .

The probability $P(E, X)$ that a meson of energy E will travel distance X without decaying is

$$\exp\left(-\int_0^X \frac{\mu c}{\tau E(x)} dx\right),$$

where μ , τ are the mass and the lifetime at rest of the meson. The only mesons we shall be concerned with in interpreting Rau's results have $E \geq 10^{10}$ ev. and for these the energy loss in penetrating the atmosphere may be neglected so that $P(E, X) = \exp(-\alpha X/E)$ where α is constant ($= \mu c/\tau$).

The total number $M(X)$ of mesons recorded at sea level

$$= K \int_{E_0} E^{-n} \exp(-\alpha X/E) dE,$$

where E_0 is the lowest energy recorded by the apparatus and K is a constant.

The change δM in M produced by a change δX in height of creation of the mesons is

$$\left[K\alpha\delta X \int_{E_0}^{\infty} E^{-(n+1)} \exp(-\alpha X/E) dE \right], \quad \dots\dots(1)$$

so that

$$\frac{\delta M}{M} = \frac{\alpha\delta X \int_{E_0}^{\infty} E^{-(n+1)} \exp(-\alpha X/E) dE}{\int_{E_0}^{\infty} E^{-n} \exp(-\alpha X/E) dE}$$

$$= \alpha\delta X \cdot F(E_0, X) \quad \text{say.}$$

or

$$\delta X = \delta M / (\alpha M F(E_0, X)). \quad \dots\dots(2)$$

For n an integer the integrands involved in $F(E_0, X)$ can be integrated by parts and the function F can be readily tabulated against X for a given E_0 .

If $\delta M/M$ is put equal to the amplitude of the semi-diurnal variation experimentally observed, then δX gives the movement of the meson-producing layer required to account for the observed variation at sea level. On the average, mesons will be produced when the primary particles have penetrated a certain mass of air, i.e. at a certain isobaric level in the atmosphere. The mean height of this level may be found by comparing the δX determined from equation (2) with Δh given in table 3.

As Rau's results represent the most accurate determination, this process was carried out using his value for $\delta M/M$ and E_0 . For $E_0 = 10^{10}$ ev. and $n=3$, $F(E_0, X)$ is a very slowly varying function of X and δX is about 160 metres for all values of X between 30 and 100 km. * Using table 3 this gives about 80 km. as the mean height of meson production.

Ehmert (1940), in a very rough calculation using a mean decay coefficient, estimated that a movement of 270 m. in the meson-producing layer was necessary to give a variation of 0.18% in meson intensity at sea level. The discrepancy between this value and δX found above seems to be largely due to the fact that Ehmert, following Euler and Heisenberg (1938), has taken different values for μ and τ (160 instead of $180 \times$ mass of electron for μ , and 2.7×10^{-6} instead of 2.1×10^{-6} sec. for τ). If Ehmert's value for δX is used however, the height of meson formation is only altered by about 8 km.

In the above calculation it has been assumed that mesons all arise at the same height, but in actual fact meson production will be spread over a considerable range in height. The effect of this spread will be to lower the height at which maximum production occurs, for the oscillation of the isobaric levels given in table 3 increases rapidly with height above 65 km., and mesons produced higher than the average have a much greater effect on the variation at sea level than those produced lower. Although the result of the simple calculation described above is not seriously modified by this consideration, a more accurate calculation can be made as shown in the next section.

* All mesons are assumed to have the same mass ($200 \times$ mass of electron) and the same lifetime at rest (2.15×10^{-6} sec.).

3.2.2 Treatment taking account of the distribution of meson production with height

Suppose that

$P(h)$ = number of primary particles per unit area at height h above sea level.

σ = cross section (assumed constant) for production of mesons by these primary particles.

$\rho(h)$ = number of nuclei per unit volume at height h .

$f(h)$ = number of mesons produced per unit length.

Then

$$f(h) = dP/dh = \sigma P(h) \rho(h). \quad (3)$$

For an isothermal atmosphere,

$$\rho(h) = \rho_0 e^{-h/H}, \quad (4)$$

where H is a constant often called the scale height; in this case

$$dP/dh = \rho_0 \sigma P(h) e^{-h/H}. \quad (5)$$

By integrating (5),

$$\ln P = -Ae^{-y} \quad \text{where} \quad A = H\rho_0\sigma; \quad y = h/H. \quad (6)$$

Hence

$$dP/dy = Ae^{-y} \exp(-Ae^{-y}) = ze^{-z}, \quad (7)$$

where

$$z = Ae^{-y} \quad \text{and} \quad \ln z = \ln A - y = -y + \text{constant}. \quad (8)$$

It follows that a plot of ze^{-z} against $\ln z$ will represent the rate of meson production, $f(y)$, as a function of y ; this is shown in figure 1. The position of sea level ($y=0$) in this diagram is determined by the value of $\ln A$ which depends only upon the cross-section—a change in cross-section displaces the curve horizontally but does not alter its shape. The fact that the same curve applies for all values of σ greatly simplifies the numerical work involved.

The curve shown in figure 1 strictly applies only for an isothermal atmosphere (H constant). The calculation of this curve is considerably more complicated when H is a varying function of height as in the true atmosphere; it has been carried out for the true atmosphere for one particular value of σ . As the curve

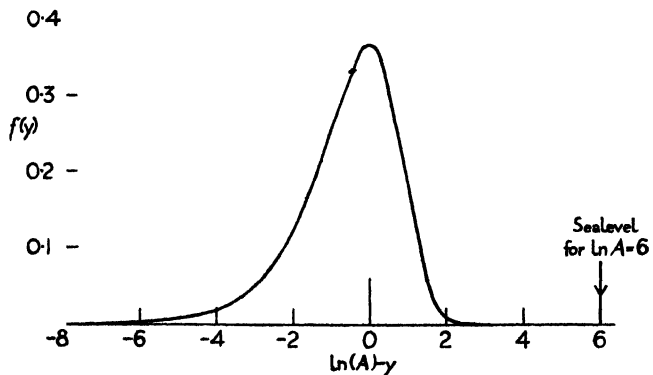


Figure 1. Rate of meson production per unit length as a function of height.

obtained in this case was very similar in shape to that for H constant and only displaced slightly in height, it has been assumed that the variation in H may be neglected for present purposes.

Figure 1 gives the rate of meson production for a production process with a particular cross-section σ . In actual fact σ will depend upon the energy of the primary particle and also upon the energy of the meson it produces, and the meson production actually occurring in the atmosphere will be the result of superposing curves corresponding to a whole range of values of σ . However, as there is no reliable theoretical or experimental knowledge of the dependence of σ on energy, it seems reasonable to assume for the entire production an average cross-section independent of energy.

Suppose the mesons have a differential energy spectrum E^{-n} at their point of origin (n will be taken equal to 3 in numerical work). Then M , the total number of mesons reaching sea level

$$\int_{h=0}^{\infty} dh f(h) \int_{E_0}^{\infty} E^{-3} \exp(-\alpha h/E) dE. \quad \dots\dots(9)$$

We require to find the change in M resulting from a movement Δh of the isobar, whose mean position is at height h above sea level. We shall assume that the rate of production of mesons remains the same at a particular isobaric level throughout the semi-diurnal atmosphere oscillation. This is justified provided the density at a given isobaric level does not vary significantly, which is the case below 90 km. at least. With this assumption

$$\delta M = \alpha \int_{h=0}^{\infty} f(h) \Delta h dh \int_{E_0}^{\infty} E^{-4} \exp(-\alpha h/E) dE, \quad \dots\dots(10)$$

so that

$$\frac{\delta M}{M} = -\alpha \frac{\int_0^{\infty} f(h) \Delta h F_1(h) dh}{\int_0^{\infty} f(h) F_2(h) dh}, \quad \dots\dots(11)$$

$$\text{where } F_1 = \int_{E_0}^{\infty} E^{-3} \exp(-\alpha h/E) dE, \quad F_2 = \int_{E_0}^{\infty} E^{-4} \exp(-\alpha h/E) dE. \quad \dots\dots(12)$$

F_1 and F_2 can be found by integration by parts and are readily tabulated as functions of h . The integrals occurring in (11) are then evaluated by numerical integration. This evaluation has been performed for three different values of σ , using the values of Δh provided by Weekes and Wilkes at latitude 50° and taking $E_0 = 10^{10}$ ev., in accordance with Rau's measurements. The results are given in table 4.

Table 4. The semi-diurnal variation in meson intensity at sea level for three different values of σ

ln A	6	8	9
A	$e^6 \sim 400$	$e^8 \sim 3000$	$e^9 \sim 8100$
Ht. of max. meson production (km.)	48	64	72
$\delta M/M$	0.0008	0.0014	0.0017

Rau's measurements give $\delta M/M = 0.0016$, and so to explain these measurements by means of the Pekeris theory the mesons have to be created mainly at a height of about 70 km.*

* Before the calculation which resulted in table 4 was carried out it was thought that the height of meson formation might be reduced to as low as 50 or 60 km. by taking account of the spread in height of meson formation, and in a brief statement on this work at a Physical Society Conference in July 1946, about 50 km. was suggested as the mean height of meson formation required to explain Rau's results (Sarabhai and Nicolson 1947).

3.3 *The high cross-section*

It is concluded above that the bulk of the mesons must arise at about 70 km. above sea level if Rau's results are to be accounted for. At 70 km. the pressure is about 0.01 cm. Hg, which is roughly $1/8000$ of atmospheric pressure. This means that primary particles have to create mesons in traversing an amount of air equivalent to only $1/8000$ of the atmosphere. The corresponding cross-section for the production process is considerably larger than seems likely if protons are the meson primaries. The maximum conceivable cross-section in this case, in air, is of the order of the geometrical cross-section of the oxygen nucleus and corresponds to a mean free path for the proton of about $1/20$ of the atmosphere, so that the cross-section derived above is over 100 times the largest reasonable value for proton primaries.

It has been suggested (Hoyle 1947) that the primary hard radiation is not composed of protons but of very heavy nuclei of atomic number Z about 80. Since the energy a particle needs to penetrate the earth's magnetic field is proportional to its charge, such a nucleus would need energy about 10^{12} ev. to reach the equator. Its fate upon impact with oxygen and nitrogen atoms at the top of the atmosphere is unknown. It is probable that collisions with air nuclei will lead to rapid disintegration of the heavy nucleus but it is uncertain whether mesons will arise directly from these disintegrations or only indirectly through the agency of protons and neutrons which will be liberated. For direct collisions involving meson interaction, the maximum cross-section for meson production by nuclei with Z about 80 cannot be much larger than about twelve times that for production by protons. (This estimate corresponds to the appropriate geometrical cross-section, on the assumption that the nuclear radius is proportional to (atomic weight)^{1/3}). This is smaller than the cross-section required to explain the $+S$ variation by a factor of more than 10. Larger cross-sections for meson production by a heavy nucleus might arise as a result of the Coulomb interaction, but it seems very unlikely that for $Z=80$ the cross-section could be as large as that needed to account for the $+S$ effect.

It is hard to see how this serious difficulty in the explanation of the $+S$ effect may be resolved. The Pekeris theory of atmospheric oscillations appears to be well-established; Weekes and Wilkes believe that the results quoted in table 3 are substantially correct up to about 60 km. On the other hand, for a semi-diurnal variation which is positively correlated with pressure, it is difficult to think of an alternative explanation to that in terms of the Pekeris theory. The first need is for more experimental evidence of a reliable kind on the existence and magnitude of this $+S$ variation. Since the pressure oscillation decreases rapidly with increase in latitude, reliable evidence is most readily obtained at the equator. One of us (Sarabhai) hopes to obtain such evidence in India shortly.

§ 4. THE $-S$ VARIATION

The experimental data suggest that, although the intensity measured under a thick absorber shows a $+S$ variation, that measured without an absorber shows a $-S$ variation. This implies that the easily absorbed radiation, consisting of the soft component and low energy mesons, is characterized by a $-S$ variation which is large enough to swamp the $+S$ variation shown by the energetic mesons.

If the soft radiation arises in the lower atmosphere as is certainly the case for the electronic component observed at sea level, it will tend to show a $-S$ variation.

on account of the ordinary mass absorption effect. However, even for the electronic component it is difficult to say what would be the magnitude of the $-S$ variation produced purely by pressure changes. The effect on electrons arising from a primary electronic component could probably be estimated fairly well from the cascade theory, but the effect on the remaining electrons which arise from mesons and which probably form about $\frac{2}{3}$ of the electrons at sea level, is extremely difficult to assess. It is likely that about equal numbers of these electrons arise by knock-on processes and by decay, and as their production is spread over various strata of the lower atmosphere, individual electrons may be influenced by different pressure changes in reaching sea level. In addition the point of origin of the mesons which produce the electrons is affected by pressure changes, again in a way depending upon altitude. In fact the position is so complicated that it seems almost impossible to calculate the resultant effect (which is known to be small) with any certainty.

An alternative approach is from the experimental measurements of the day-to-day variation of intensity and the corresponding pressure variation, i.e. determination of what is usually called the barometric coefficient. Here again no definite conclusion may be reached, for as already mentioned, it is uncertain how much of the observed barometric effect is a true mass-absorption effect and how much is produced by other factors, such as movement of meson-producing layers, which may be differently related to semi-diurnal and to day-to-day pressure changes. Duperier finds a barometric coefficient of 0.35% per mm. Hg and deduces that about 60% of this is a mass absorption effect and that the remainder is caused by movement of meson-producing layers; this result is dependent upon mesons being created at about 16 km. which may not be the case. It does appear fairly certain, however, that the $-S$ effect observed by Duperier with no absorber is too large to arise purely as a mass-absorption effect from pressure changes.

Although the explanation of the $-S$ effect is not clear, the suggestion made by Duperier (1945) that the main factor responsible is the atmospheric oscillation predicted by Pekeris is untenable when the magnitude of this oscillation is considered. A negative S effect can only be produced by the Pekeris oscillation if mesons arise below 30 km. but the maximum displacement of isobars below 30 km. is only about 3 metres at latitude 50° and this movement is insufficient to produce an observable cosmic-ray effect. Duperier appears to be considering the vertical movement of individual particles, which indeed may be as large as 20 metres, as he states, but which is irrelevant in a consideration of the height of meson production, where the important quantity is the vertical movement of isobaric level.

Further confusion on this point has arisen from recent publications by two workers in Ceylon, which apparently confirm Duperier's conclusion that the $-S$ variation can be accounted for by the Pekeris theory (Mailvaganam 1946, Kidnapillai and Mailvaganam 1946). It has already been mentioned (§ 3.1) that the results obtained by these workers for the movement of isobars are incorrect. Apart from this, their final result of 0.14% for the amplitude of the $-S$ effect seems to apply at $\lambda = 0$, whereas Duperier's value of 0.18%, with which they compare it, applies at $\lambda = 50^\circ$ and if produced by pressure oscillations would be four times as large at $\lambda = 0$.

§ 5. THE DIURNAL VARIATION

From § 2.1 it will be seen that although the D variation at high latitudes, recorded by different observers using similar apparatus, usually has an amplitude

about 0.2% the phase varies appreciably even at stations having the same latitude. For example, the results of Hess and Graziadei (1936) and of Forbush (1937), which both appear to be well established show a discrepancy of several hours in phase when allowance is made for the fact that with no external temperature correction the maximum observed by Hess and Graziadei would probably occur about four hours earlier. Duperier (1941) observes a maximum as late as 17^h, but here the discrepancy may arise partly from the difference in thickness of absorber used. Doan (1936) whose results are fairly convincing, places the maximum at 9^h. In fact all that can be concluded in general about the time of maximum intensity seems to be that it occurs during the daytime rather than during the night.

It is possible that the recordings made with ionization chambers, even when shielded with 10 cm. Pb, are affected by fluctuations in the gamma-ray content of the atmosphere. The latter, in some localities at least, appears to show a daily variation with maximum about noon. Forbush (1937) estimates that this factor might cause as much as one third of the D variation recorded by his apparatus and suggests that it might produce quite different distortion of the true cosmic-ray D variation in experiments performed in different localities. However, it appears very unlikely that the wide phase variation recorded in the experiments mentioned above can be entirely due to this effect. The results of Alfvén and Malmfors (1943) recorded in Stockholm with Geiger counter apparatus and supported by those of Kolhorster (1941) in Berlin, indicate that the D effect is appreciably different in different directions with the same zenith angle, and in view of this it does not seem improbable that the true D variation at a particular zenith angle will be different in different localities.

The variation in character of the experimentally observed D effect has been ignored by most workers seeking an explanation of the effect; following the early results of Hess they usually assume that the maximum occurs at midday and try to account for this. For instance, Vallarta and Godart (1939) suggest that such a variation can be accounted for at high latitudes by a heliomagnetic field and at low latitudes by fluctuations in the geomagnetic field. It is clear, however, that the D effect indicated by a wider survey of experimental data could not be mainly produced by either a heliomagnetic field or by geomagnetic fluctuations since it appears to show an appreciable variation in phase at different stations at the same latitude. Malmfors (1945) shows that such explanations are also incompatible with the results found by Alfvén and himself (1943) in Stockholm. These results are particularly interesting as they also eliminate meteorological changes as the main cause of the D variation. Malmfors suggests that occasional small disturbances of the isotropy of the primary radiation in remote space may be responsible for the Stockholm results.

Before the cause of the D effect can be considered further, more experiments at various latitudes and longitudes are needed to establish the variation of the phase of the effect under a given absorber with geographical situation and also the variation at a given station of phase and amplitude with thickness of absorber and with angle θ .

ACKNOWLEDGMENTS

The authors wish to acknowledge their indebtedness to Mr. K. Weekes and Dr. M. V. Wilkes of Cambridge University, for furnishing the data quoted in table 3 and also for many helpful discussions on the subject of atmospheric oscillations. In addition they would like to thank Professor P. M. S. Blackett and

Dr. L. Jánossy of Manchester University for stimulating discussions on the cosmic-ray aspect of the problem.

REFERENCES

- ALFVÉN and MALMFORS, 1943, *Ark. Mat. Astr. Fys.*, **29**, No. 24.
BARNOTHY and FORRO, 1936, *Z. Phys.*, **100**, 742 ; 1937, *Ibid.*, **104**, 534 ; 1939, *Phys. Rev.*, **55**, 868.
BUDDEN, RATCLIFFE and WILKES, 1939, *Proc. Roy. Soc. A*, **171**, 188.
CHAPMAN, 1919, *Philos. Trans.*, **218**, 1 ; 1924, *Quart. J.R. Met. Soc.*, **50**, 165.
DOAN, 1936, *Phys. Rev.*, **49**, 107.
DUPERIER, 1944, *Terr. Magn. Atmos. Elect.*, **49**, 1 ; 1945, *Proc. Phys. Soc.*, **57**, 468.
EHMERT, 1940, *Naturwissenschaften*, **28**, 28.
EULER and HEISENBERG, 1938, *Erbegn. Exakt. Naturw.*, **17**, 1.
FORBUSH, 1937, *Terr. Magn. Atmos. Elect.*, **42**, 1.
HESS and GRAZIADEI, 1936, *Terr. Magn. Atmos. Elect.*, **41**, 9.
HOYLE, 1947, *Roy. Ast. Soc.* (in press).
KIDNAPILLAI and MAILVAGANAM, 1946, *Phys. Rev.*, **70**, 94.
KOLHORSTER, 1941, *Phys. Z.*, **42**, 55.
MAILVAGANAM, 1946, *Proc. Phys. Soc.*, **58**, 468.
MALMFORS, 1945, *Ark. Mat. Astr. Fys.*, **32 A**, No. 8.
PEKERIS, 1937, *Proc. Roy. Soc. A*, **158**, 650 ; 1939, *Ibid.*, **171**, 634.
RAU, 1939, *Z. Phys.*, **114**, 265 ; 1940, *Ibid.*, **116**, 105.
REGENER and RAU, 1939, *Naturwissenschaften*, **27**, 803.
SARABHAI, 1945, *Proc. Indian Acad. Sci., Bangalore*, **A**, **21**, 66.
SARABHAI and NICOLSON, 1947, *Phys. Soc. Intern. Conf.: I—Fundamental Particles*, 68.
SCHONLAND, DELATISKY and GASKELL, 1937, *Terr. Magn. Atmos. Elect.*, **42**, 137.
STEINMAURER, 1935, *Beitr. Geophys.*, **45**, 148.
TAYLOR, 1929, *Proc. Roy. Soc. A*, **126**, 169 ; 1932, *Mem. R. Met. Soc.*, **4**, No. 35 ; 1936, *Proc. Roy. Soc. A*, **156**, 318.
THOMPSON, 1938, *Phys. Rev.*, **54**, 93.
WHIPPLE, 1934, *Quart. J.R. Met. Soc.*, **60**, 51 ; 1935, *Ibid.*, **61**, 285.
VALLARTA and GODART, 1939, *Rev. Mod. Phys.*, **11**, 180.

The Neutrons Emitted in the Disintegration of Nitrogen by Deuterons

BY W. M. GIBSON AND D. L. LIVESEY

Cavendish Laboratory, Cambridge

MS. received 18 December 1947

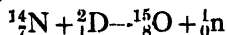
ABSTRACT. The photographic plate method has been used in a study of the neutrons produced in the disintegration of nitrogen by deuterons.

A single group of neutrons, corresponding to an energy release of (5.15 ± 0.10) mev., was detected, but the existence of the low-energy neutron group reported by Stephens, Djanab and Bonner (1937) was not confirmed. The new results show that, if this group does exist, its intensity is not more than 15% of that of the principal neutron group, and that no excited levels of the ^{16}O nucleus less than 5 mev. above the ground state are effective in this reaction.

The neutron flux from the $\text{N} + \text{D}$ reaction was compared with that of $\text{D} + \text{D}$ neutrons produced by disintegration of deuterium present in the surface layers of the target employed. The data obtained were used in determinations of the angular distributions of the $\text{D} + \text{D}$ and $\text{N} + \text{D}$ neutron fluxes, and in tests of the reliability of the photographic plate method for relative flux measurements.

§ 1. INTRODUCTION

THE neutrons produced in the disintegration of nitrogen by deuterons according to the scheme :



have been investigated by Stephens, Djanab and Bonner (1937), who used an expansion chamber and detected three groups of neutrons emitted from nitrogen targets bombarded by deuterons. They attributed two of the groups to the above reaction, and the effective Q values were found to be (1.1 ± 0.2) mev. and (5.1 ± 0.2) mev. respectively. The third group was ascribed to the disintegration by deuterons of deuterium present in the surface layers of the targets employed, and the Q value of this group (3.3 mev.) was in agreement with the known energy release of the $\text{D} + \text{D}$ reaction.

Further investigation of the $\text{N} + \text{D}$ reaction is desirable in order to confirm the existence of an excited level in the ${}^{15}_8\text{O}$ nucleus, which is indicated by the presence of the low-energy group in the $\text{N} + \text{D}$ neutron spectrum. In addition, the 5 mev. neutron group may be useful in experiments where a neutron source of this energy is required, and for this purpose the characteristics of the neutron spectrum must be determined at various angles to the direction of the deuteron beam.

The use of photographic plates in the determination of neutron energies has been described by Powell (1940, 1942) and Richards (1941 a), who have studied the distributions in energy of the neutrons emitted in the bombardment of a number of light elements by deuterons. Their results were in general agreement with those obtained by the expansion chamber method. The photographic plate technique has the advantage that plates may be exposed simultaneously at different angles to the direction of the primary beam, so that information may be obtained concerning the angular distribution of the neutron yield (Richards 1941 b, Livesey and Wilkinson 1948). The photographic technique has been improved considerably as a result of the introduction by Ilford Ltd. of concentrated emulsions (Powell, Occhialini, Livesey and Chilton 1946), with which greatly increased accuracy of track measurements is attainable, and at present it is the most convenient and precise method for the determination of neutron spectra. This method was therefore adopted for a more detailed investigation of the neutrons emitted in the disintegration of nitrogen by deuterons.

§ 2. EXPERIMENTAL METHOD

A thick target of aluminium nitride was bombarded with a beam of deuterons accelerated through 930 kilovolts by the Cavendish Laboratory high tension equipment.

Four light-tight boxes containing photographic plates of the Ilford C.2 type with emulsions 100 microns thick were set up around the target so that the plates recorded neutrons emitted at 0° , 30° , 90° and 150° respectively.

In order to obtain reasonably good geometrical conditions with the relatively large neutron source, the plates were placed 15 cm. from it and were irradiated for 3 hours with an average beam current of $80 \mu\text{a}$. This arrangement gave a maximum spread of $\pm 3^\circ$ in the directions of the neutrons incident upon the plates. A stop could not usefully be employed to reduce the size of the $\text{N} + \text{D}$ neutron source, as the stop itself acted as a source of $\text{D} + \text{D}$ neutrons.

A second series of plates was irradiated with neutrons emitted from a pure aluminium target bombarded with deuterons, in order to estimate the contribution of the aluminium to the spectrum of neutrons from the aluminium nitride target. The conditions of exposure were the same as for the nitride target, except that the target surface was cooled less efficiently and was therefore maintained at a higher temperature by the beam.

After the plates had been processed, detailed microscope measurements were carried out on the proton tracks occurring in the plates by two observers, using a Leitz microscope fitted with a 1/12 inch oil-immersion objective which, in conjunction with twin eyepieces of power $\times 10$, gave an overall magnification of $\times 1200$. Traverses were made across each plate in a direction parallel to that of the incident neutrons, and all proton tracks were recorded which originated in the area covered by the traverse, which were completely contained within the emulsion, and which made angles of less than 20° with the neutron direction. Between 300 and 500 tracks were recorded in each plate. The area examined was calculated from the total length of the traverse; it was approximately 20 mm^2 in all cases.

For each selected track the total length a projected on the focal plane of the microscope was measured, and the initial direction of the proton was derived from measurements of the angle of scattering projected on the focal plane α and the initial angle of dip ψ . In order to obtain ψ it was necessary to correct the distances measured in the vertical plane for the effects of shrinkage of the emulsion caused by fixation. The correction consisted in multiplying all vertical distances by a factor s , which was assumed to be equal to the ratio of the initial to the final thickness of the emulsion. A series of experimental determinations of the shrinkage factor indicated $s = 2.0 \pm 0.2$, and this value was used in all calculations of the angle ψ . The actual angle of scattering γ for each proton track was calculated, to the nearest degree, from the relation $\sec \gamma = \sec \alpha \sec \psi$.

The proton range R was calculated from the equation $R = a \sec \psi$. This formula applies strictly only to tracks which are straight; in fact, some of the tracks were deflected from straight lines by nuclear collisions, but these deflections did not produce appreciable errors in the estimates of R and γ . The proton energy E_p was derived from the range-energy relation obtained for the B.1 type of emulsion by Lattes, Fowler and Cuer (1947). A comparison of the properties of the B.1 and C.2 emulsions has shown that the stopping-powers of the two types for α -particles are very nearly identical over a wide range of energies. It is therefore reasonable to assume that the stopping-powers for protons are the same. The neutron energy E_n is related to the proton energy E_p by the formula $E_n = E_p \sec^2 \gamma$. The quantity E_n was calculated for each individual track recorded.

§ 3. NEUTRON SPECTRA

The energy distributions of the neutrons emitted from the aluminium nitride target at the four angles of observation ($\phi = 0^\circ, 30^\circ, 90^\circ$ and 150°), and from the aluminium target at $\phi = 0^\circ$, were derived from the numbers of tracks recorded in successive intervals of 0.1 MeV in energy E_n . The data were divided, according to the magnitude of the scattering angle γ , into three groups defined by the conditions :

- (i) $0^\circ \leq \gamma \leq 9^\circ$; (ii) $10^\circ \leq \gamma \leq 14^\circ$; (iii) $15^\circ \leq \gamma \leq 19^\circ$.

A typical track distribution is shown in figure 1, which refers to tracks measured in the 150° nitride plate.

It was found that, in the energy distributions obtained for $15^\circ \leq \gamma \leq 19^\circ$, the D+D and N+D neutron groups were broader than those occurring in the distributions obtained for low values of γ . This effect was due to small errors in definition of the directions of the incident neutrons, caused by the use of a comparatively large neutron source. This "angular straggling" is always more pronounced in the case of protons scattered at large angles, and measurements on these tracks are less reliable for the determination of neutron energies. Despite the straggling effects the principal neutron groups were clearly resolved in the distributions obtained for $\gamma < 20^\circ$, and no

difficulty was experienced in estimating the number of proton tracks which fell within each group. The data for $0^\circ \leq \gamma < 9^\circ$ were used in all energy determinations, since the effects of angular straggling were very small in these cases.

Since the total number of tracks measured in each plate was comparatively small, statistical variations in the number of tracks at each energy were considerable. The observed distributions were therefore smoothed by the addition to each ordinate on the energy scale of the mean of the two neighbouring ordinates. In order to obtain the actual neutron energy distributions, the smoothed experimental distributions were corrected, firstly for the variation of the total neutron-proton scattering cross-section with energy, the formulae of Kittel and Breit (1939) being used, and secondly for the escape of protons from the emulsions. The correction for the latter effect was calculated geometrically on the assumption that the proton range varies with energy according to the Geiger rule, which, for a fixed neutron energy E_n , yields the following relation between the proton range R and the scattering angle γ : $R = R_0 \cos^3 \gamma$ where R_0 is the range of a proton with energy E_n . The fraction f of the total number of protons, scattered in the range $\gamma_1 \leq \gamma \leq \gamma_2$, which is recorded in an emulsion of thickness $2d$, is given by $f = 1 -$

$$\frac{R_0}{8\pi d} \frac{\{(\gamma_2 - \gamma_1) + \frac{1}{4}(\sin 2\gamma_2 - \sin 2\gamma_1) - \frac{1}{4}(\sin 4\gamma_2 - \sin 4\gamma_1) - \frac{1}{12}(\sin 6\gamma_2 - \sin 6\gamma_1)\}}{(\sin^2 \gamma_2 - \sin^2 \gamma_1)},$$

provided that $R_0 \cos^3 \gamma_2 \sin \gamma_2 \leq 2d$. Although the Geiger rule does not accurately represent the range-energy relation for protons, the correction factor is not

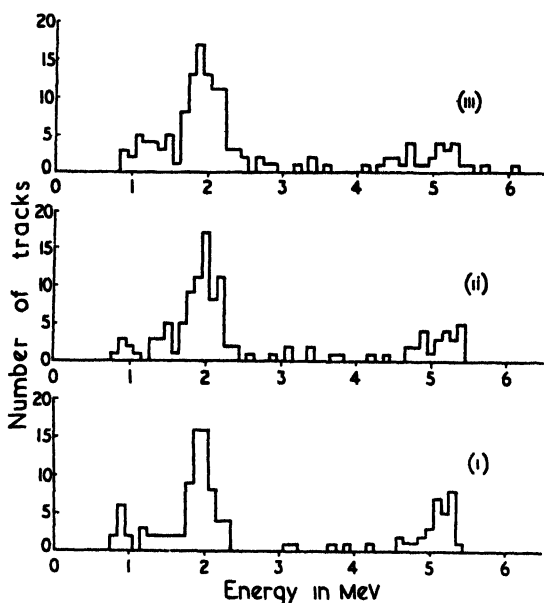
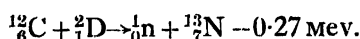


Figure 1. Number of tracks plotted against neutron energy for 150° plate:

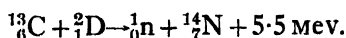
(i) $0^\circ \leq \gamma \leq 9^\circ$; (ii) $10^\circ \leq \gamma \leq 14^\circ$; (iii) $15^\circ \leq \gamma \leq 19^\circ$.

appreciably affected by the departures from it which actually occur. The calculated factor is plotted as a function of the neutron energy in figure 2 for the three ranges of γ used in the experiment.

The corrected neutron spectra for the aluminium nitride target at four angles of observation and for the aluminium target at $\phi = 0^\circ$, are shown in figure 3. The four nitride spectra show two principal neutron groups, the mean energy of the one varying between 1.9 Mev. at 150° and 3.5 Mev. at 0° , and that of the other lying between 5 Mev. and 6 Mev. at all angles. In addition, a third group below 1 Mev. was detected in the 0° spectrum, the investigation of which was extended to lower energies than in the other cases. This group is ascribed to the disintegration of carbon present on the target surface, according to the reaction :



A similar group was found in the 0° aluminium spectrum. The occurrence of the $^{13}\text{C} + \text{D}$ neutron group indicates that high-energy neutrons were emitted also as a result of the disintegration of the isotope ^{13}C :



It has been shown by Bonner and Brubaker (1936) that the abundance of the $^{13}\text{C} + \text{D}$ neutron group is only 1% of that of the $^{12}\text{C} + \text{D}$ group. It is clear, therefore, that the $^{13}\text{C} + \text{D}$ reaction cannot account for more than a small proportion of the 5–6 Mev. neutrons occurring in the nitride spectra, and these may be identified with the $\text{N} + \text{D}$ neutron group investigated by Stephens, Djanab and Bonner (1937). Nevertheless, a small neutron component was found near 6 Mev. in the 0° aluminium spectrum, and this may be ascribed to the $^{13}\text{C} + \text{D}$ reaction.

The remaining principal group in the nitride spectra was produced by the disintegration of deuterium present in the surface layers of the target, as may be seen by comparison with the $\text{D} + \text{D}$ neutron spectra obtained by Livesey and Wilkinson (1948), who deduced a value of 3.23 ± 0.02 Mev. for the energy release of the reaction. The shape of the $\text{D} + \text{D}$ neutron group in the 0° aluminium spectrum is different from that of the $\text{D} + \text{D}$ group in the 0° nitride spectrum, and this is probably due to the high temperature at which the aluminium target was maintained : under these conditions the material closest to the surface of the target retained only a very small amount of deuterium, and so the yield of neutrons of maximum energy was low.

Between 0.8 Mev. and 2.0 Mev. no clearly defined group was detected in the nitride spectra ; this result disagrees with those of Stephens, Djanab and Bonner.

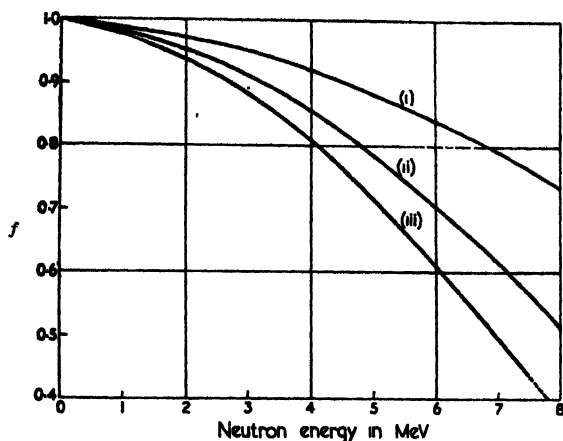


Figure 2. Escape fraction f as a function of neutron energy :
(i) $0 \leq \gamma \leq 9^\circ$; (ii) $10^\circ \leq \gamma \leq 14^\circ$; (iii) $15 \leq \gamma \leq 19^\circ$.

A comparison of the nitride and aluminium spectra at 0° shows that the distributions of neutron energy between 0.8 Mev. and 2.0 Mev. are very nearly the same in the two cases. The ratio of the neutron fluxes in the two spectra (Al flux/AlN flux) was determined directly, and the values obtained for the energy ranges 1.0–1.5 Mev. and 1.5–2.0 Mev. were 1.6 ± 0.2 and 1.4 ± 0.2 respectively.

The fact that the aluminium target produced more neutrons in the energy range 1.0–2.0 Mev. than the aluminium nitride target shows that nitrogen did not contribute appreciably to this group of neutrons. It is known that the carbon isotope ^{13}C emits neutrons of this energy when bombarded with deuterons, but the abundance of the $^{13}\text{C} + \text{D}$ group, as calculated from the results of Bennett, Bonner, Hudspeth, Richards and Watt (1941), is too small to account for the

observed effect. It is possible that the aluminium present in both targets produced most of the neutrons in this energy range. If this hypothesis is correct, the ratio of neutron fluxes, as calculated from the chemical compositions of the targets, should be 1.5, which is in agreement with the experimental value. An upper limit may therefore be set to the abundance of the low-energy $\text{N} + \text{D}$ group reported by Stephens, Djanab and Bonner (1937). On the assumption that the maximum variation in the neutron flux ratio is three times the probable

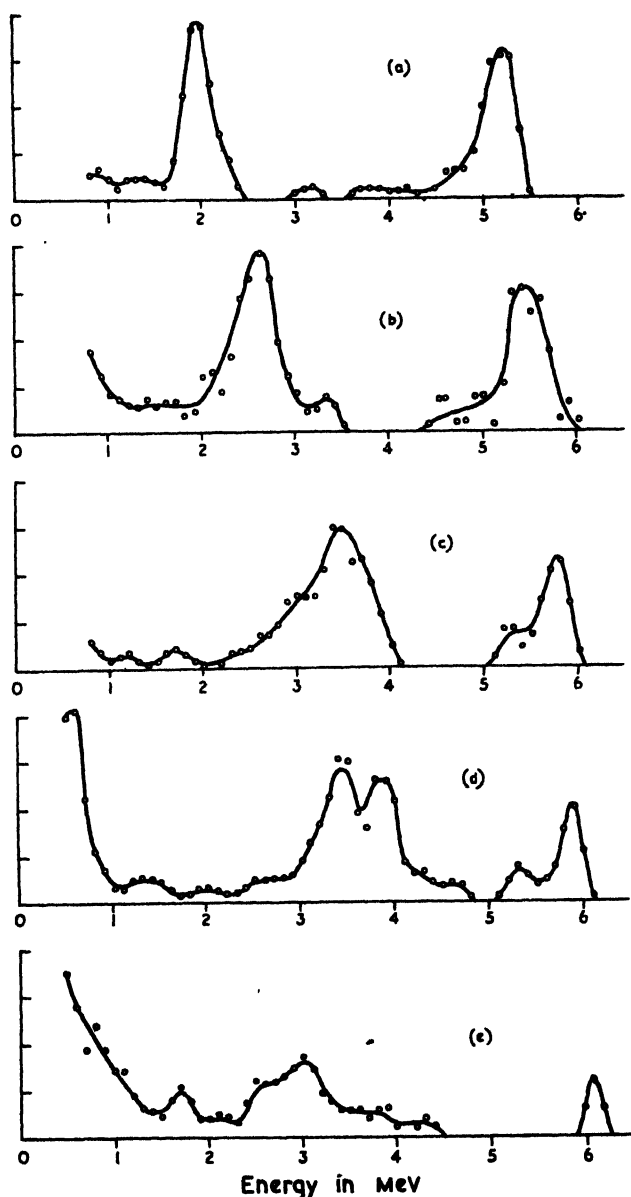


Figure 3. Corrected neutron spectra, with ordinates in arbitrary units: (a) AlN target at 150° , (b) AlN target at 90° , (c) AlN target at 30° , (d) AlN target at 0° , (e) Al target at 0° .

error, this upper limit is found to be 15% of the abundance of the high-energy N + D group at the same angle, namely 0°. This same conclusion is valid for all four angles of observation, and it disagrees entirely with the results of Stephens, Djanab and Bonner, who estimated the abundance of the low-energy group to be 150% compared with that of the high-energy group, at an angle of 90° to the direction of the deuteron beam.

The high-energy N + D neutron group is well-defined and is clearly resolved from other components in all the neutron spectra investigated. The mean thick-target energy varies between 5.1 mev. at 150° and 5.7 mev. at 0° when the bombarding energy is 0.93 mev.

§ 4. THE ENERGY RELEASE OF THE N + D REACTION

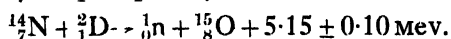
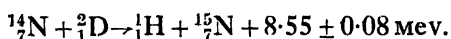
In order to obtain the energy release of the N + D reaction, it was necessary to extract, from the observed thick-target spectrum for each of the four angles of observation, the corresponding thin-target energy. A careful examination of the spectra of figure 3 showed that the effects of both angular straggling and range straggling of the proton tracks could be represented analytically by the superposition of a generalized Gaussian error function on the actual thick-target neutron spectrum. A method described by Livesey and Wilkinson (1948) was therefore employed to derive the thin-target energy for each spectrum. The N + D spectrum at each angle was re-plotted, with the same corrections and smoothing procedure as those previously described, but with the neutron energy specified at 0.05 mev. intervals, and the integrated spectrum was constructed in each case. The correction required for estimating the thin-target energy (E_0) from the value given by extrapolation of the straight part of the integrated spectrum was obtained by comparing the shape of the experimental spectrum with those of a set of theoretical curves calculated for different values of the Gaussian straggling parameter. The extrapolated energies, the derived thin-target energies, and the corresponding values for the energy release Q are included in table 1.

Table 1. Extrapolated energies, thin-target energies and calculated Q values in mev.

Angle ϕ°	0	30	90	150
Extrapolated energy	6.00	5.94	5.82	5.39
Correction	-0.08	-0.10	-0.13	-0.10
Thin-target energy, E_0	5.92	5.84	5.69	5.29
Q value	5.07	5.05	5.27	5.20

These figures were checked by independent determinations of E_0 from the spectra corresponding to angles of scattering γ between 10° and 14°, and no discrepancies were found. If the four estimates of the Q value are of equal weight, the mean value is 5.15 mev., and the probable error, derived from the above results only, is ± 0.06 mev. The uncertainties in the range-energy relation of Lattes, Fowler and Cier (1947) must also be taken into account; the probable error in their data for energies between 5 mev. and 6 mev. is ± 0.07 mev. In addition, the error of $\pm 10\%$ in the shrinkage factor of the emulsions leads to a probable error of ± 0.03 mev. in the final Q value. These errors in combination give a total probable error of ± 0.10 mev., and the energy release is therefore quoted as $Q = 5.15 \pm 0.10$ mev.

This result agrees with the value of 5.1 ± 0.2 mev. derived by Stephens, Djanab and Bonner (1937), and it may also be checked by calculation of the maximum energy of the positrons emitted by the oxygen isotope, ^{15}O . The (d, p) reaction in nitrogen has been studied by Cockcroft and Lewis (1936) and by Holloway and Moore (1940), and the most accurate results give a value of 8.55 ± 0.08 mev. for the energy release. Comparison of the two reactions :



yields the relation $^{15}_8\text{O} - ^{15}_7\text{N} = - (^1_0\text{n} - ^1_1\text{H}) + 3.40 \pm 0.13$ mev. The neutron-hydrogen mass difference is equivalent to 0.76 mev., according to Mattauch (1942) and Stephens (1947), hence the difference in the masses of $^{15}_8\text{O}$ and $^{15}_7\text{N}$ is equivalent to 2.64 ± 0.13 mev. This exceeds the maximum positron energy from ^{15}O by an amount equivalent to the mass of two electrons, that is, 1.02 mev., provided that the mass of the neutrino is negligible. The maximum positron energy is therefore 1.62 ± 0.13 mev., and this result agrees with the experimental value of 1.72 mev. deduced from the data of Konopinski (1943).

The results shown in figure 3 did not confirm the existence of the excited state of ^{15}O which was indicated by the results of Stephens, Djanab and Bonner (1937). A detailed examination of the nitride spectra, and especially of the 150° spectrum, leads to the conclusion that the energy release of the N + D reaction is single-valued and that no excited levels of the ^{15}O nucleus less than 5 mev. above the ground state are operative in this reaction.

§ 5. ANGULAR DISTRIBUTIONS OF NEUTRON YIELD

In making relative measurements of neutron flux by the photographic plate method, it is assumed that the number of proton tracks occurring per unit area within a defined solid angle, is, after correction for the escape effect, proportional to the product of the neutron flux and the total neutron-proton scattering cross-section. This assumption is valid if the angular distribution of neutron-proton scattering is independent of the neutron energy in the range investigated, and if the concentration of hydrogen in the emulsion is constant. That the latter condition was satisfied in this experiment was shown by the fact that no abnormally large variations in the abundance of proton tracks were detected in examining different parts of the plates. The angular distribution of neutron-proton scattering has been determined experimentally by Dee and Gilbert (1937), by Bonner (1937) and by Champion and Powell (1944) for neutron energies below 10 mev., and it is generally accepted that the distribution in the centre-of-mass coordinate system is effectively isotropic within this energy range.

The reliability of the method of observation and of the escape correction used in this experiment may be checked by comparison of the observed angular distribution of proton tracks with that corresponding to isotropic scattering in the centre-of-mass coordinates. Provided that the effects of angular straggling may be neglected, the number of proton tracks recorded in the range of angles defined by $\gamma_1 \leq \gamma \leq \gamma_2$ should be proportional to the integral

$$\int_{\gamma_1}^{\gamma_2} 2 \cos \gamma \sin \gamma d\gamma = (\sin^2 \gamma_2 - \sin^2 \gamma_1).$$

The experimental data were tested by dividing the total numbers of proton tracks in the D + D and N + D groups, corrected for escape, by the function

($\sin^2 \gamma_2 - \sin^2 \gamma_1$), to see whether the same quantity was obtained for the different angular ranges. The results are shown in table 2.

Table 2. Ratio of experimental and theoretical proton abundances, in arbitrary units

Spectrum	$0^\circ \leq \gamma \leq 9^\circ$	$10^\circ \leq \gamma \leq 14^\circ$	$15^\circ \leq \gamma \leq 19^\circ$
0°	35 ± 3	41 ± 3	31 ± 3
30°	33 ± 3	24 ± 2	22 ± 2
90°	18 ± 2	21 ± 2	17 ± 2
150°	24 ± 2	28 ± 2	25 ± 2

The 0° , 90° and 150° results are internally consistent, but the 30° figures show possible deviations from isotropic scattering. These deviations are probably not significant, but they may be regarded as an indication that the 30° results are less reliable than the others.

In the four neutron spectra obtained with the nitride target, the N + D and D + D neutron groups were everywhere resolved, and the number of neutrons detected in each group was readily estimated. In order to test more stringently the accuracy of the escape correction at the higher neutron energies, and also to detect whether there was any tendency on the part of the observers to count too many long tracks at high scattering angles, the ratio of the D + D flux to the N + D flux was determined for the three ranges of the scattering angle γ . In the calculation of this ratio, the mean scattering cross-sections for the two neutron groups were derived from the data of Kittel and Breit (1939). The results are shown in table 3.

Table 3. Ratio of D + D and N + D neutron fluxes

Spectrum	$0^\circ \leq \gamma \leq 9^\circ$	$10^\circ \leq \gamma \leq 14^\circ$	$15^\circ \leq \gamma \leq 19^\circ$	Mean Ratio
0°	4.1 ± 0.9	4.8 ± 0.8	4.4 ± 0.8	4.4 ± 0.5
30°	3.4 ± 0.6	4.3 ± 1.0	5.9 ± 1.3	4.4 ± 0.5
90°	1.7 ± 0.3	1.8 ± 0.3	1.3 ± 0.2	1.50 ± 0.13
150°	1.8 ± 0.2	2.5 ± 0.3	2.5 ± 0.3	2.12 ± 0.18

The internal consistency of these results shows that the method of observation and of correction for escape did not give rise to serious errors.

The consistency tests already described indicate that any errors occurring when the photographic plate method is used for comparing neutron fluxes can be reduced to low values by the methods used in this experiment. The data were used to determine the angular distribution of the D + D neutron flux in order to compare the results obtained with those of Allen, Livesey and Wilkinson (1948), who determined the angular distribution of neutrons emitted from a thick deuterium target bombarded by deuterons of energy 0.93 MeV. The angular distribution was obtained by estimating the neutron flux at each angle relative to that at 0° , and the results are shown in table 4.

Table 4. Angular distribution of D + D neutron flux

Angle ϕ°	$0^\circ \leq \gamma \leq 9^\circ$	$10^\circ \leq \gamma \leq 14^\circ$	$15^\circ \leq \gamma \leq 19^\circ$	$0^\circ \leq \gamma \leq 19^\circ$
0	1.00 ± 0.08	1.00 ± 0.07	1.00 ± 0.07	1.00 ± 0.05
30	1.15 ± 0.12	0.76 ± 0.07	0.90 ± 0.09	0.90 ± 0.05
90	0.29 ± 0.04	0.28 ± 0.03	0.26 ± 0.03	0.27 ± 0.02
150	0.62 ± 0.06	0.45 ± 0.06	0.54 ± 0.06	0.52 ± 0.04

The results for the angle range $0^\circ \leq \gamma \leq 19^\circ$ may be compared with the experimental curve of Allen, Livesey and Wilkinson in figure 4.

In general the agreement between the two sets of results is close.

The above results show that it is possible to make reliable determinations of angular distributions of neutron flux by the photographic plate method. The angular distribution of the $N + D$ neutron flux was calculated from the data referring to the entire range of scattering angles between 0° and 19° ; for ϕ equal to 0° , 30° , 90° , 150° the values obtained were 1.00 ± 0.10 , 0.91 ± 0.10 , 0.90 ± 0.08 , 1.37 ± 0.15 respectively.

The 150° figure is somewhat higher than the others, but the statistical accuracy is lower than that in the $D + D$ distribution, and the deviation from an isotropic distribution of the neutron yield is hardly significant.

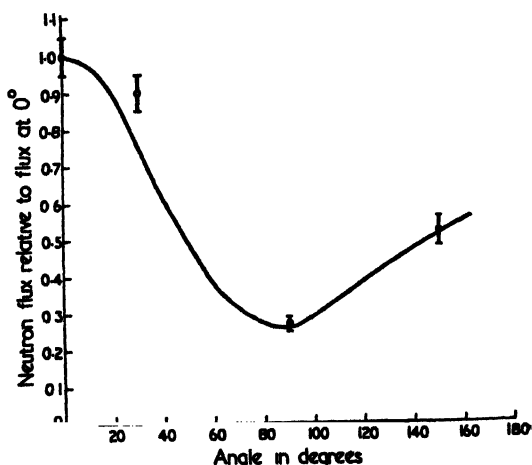


Figure 4. Angular distribution of $D + D$ neutron flux from thick target:

Experimental curve of Allen, Livesey and Wilkinson, with 3 photographic plate results for the angle range $0 \leq \gamma \leq 19^\circ$.

ACKNOWLEDGMENT

The authors wish to express their indebtedness to Dr. W. E. Burcham for helpful criticism and advice, and to the Department of Scientific and Industrial Research for financial assistance.

REFERENCES

- ALLEN, K. W., LIVESY, D. L., and WILKINSON, D. H., 1948, in course of preparation.
 BENNETT, W. E., BONNER, T. W., HUDSPETH, E., RICHARDS, H. T., and WATT, B. E., 1941, *Phys. Rev.*, **59**, 781.
 BONNER, T. W., 1937, *Phys. Rev.*, **52**, 685.
 BONNER, T. W., and BRUBAKER, W. M., 1936, *Phys. Rev.*, **50**, 308.
 CHAMPION, F. C., and POWELL, C. F., 1944, *Proc. Roy. Soc. A*, **183**, 64.
 COCKCROFT, J. D., and LEWIS, W. B., 1936, *Proc. Roy. Soc. A*, **154**, 261.
 DEE, P. I., and GILBERT, C. W., 1937, *Proc. Roy. Soc. A*, **163**, 265.
 HOLLOWAY, M. G., and MOORE, B. L., 1940, *Phys. Rev.*, **58**, 847.
 KITTEL, C., and BREIT, G., 1939, *Phys. Rev.*, **56**, 744.
 KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
 LATTES, C. M. G., FOWLER, P. H., and CUER, P., 1947, *Nature, Lond.*, **159**, 301.
 LIVESY, D. L., and WILKINSON, D. H., 1948, in course of preparation.
 MATTAUCH, J., 1942, *Kernphysikalische Tabellen* (Berlin: J. Springer).
 POWELL, C. F., 1940, *Nature, Lond.*, **145**, 155; 1942, *Proc. Roy. Soc. A*, **181**, 344.
 POWELL, C. F., OCCHIALINI, G. P. S., LIVESY, D. L., and CHILTON, L. V., 1946, *J. Sci. Instrum.*, **23**, 102.
 RICHARDS, H. T., 1941 a, *Phys. Rev.*, **59**, 796; 1941 b, *Ibid.*, **60**, 167.
 STEPHENS, W. E., 1947, *Rev. Mod. Phys.*, **19**, 19.
 STEPHENS, W. E., DJANAB, K., and BONNER, T. W., 1937, *Phys. Rev.*, **52**, 1079.

Dissociation Energy of the NO Molecule*

By L. GERÖ AND R. SCHMID

Budapest

MS. communicated by J. G. Valatin after the deaths of both authors ; received 25 June 1947

ABSTRACT. During a new investigation of the NO band spectrum many bands of the β and γ system were photographed and analysed, and the rotational constants evaluated. Intensity drops and perturbations were found and two predissociation limits at 53800 and 50500 cm^{-1} were established. In this way the dissociation energy of NO is fixed at 4.29 eV. The $x^2\Pi$ ground state of NO dissociates in the $N(^2D)+O(^3P)$ atomic term combination. The new dissociation scheme is strongly supported by the results of photochemical experiments.

IN the course of a new investigation of the band spectrum of the NO molecule, many bands of the β , γ and ϵ system were photographed with great intensity and resolving power. The rotational analysis of the ϵ system was given in a preceding paper (Gerö, Schmid and Szily 1944); after remeasuring the whole of the NO spectrum between 1950 and 2700 Å. some new data for the β and γ bands can also be given.

Light source, spectroscope and experimental details were described in the paper mentioned above. The rotational analysis of the following γ bands ($A^2\Sigma-x^2\Pi$) was carried out up to quite large rotational quantum numbers: (0,0), (0,1), (0,2), (0,3), (0,4), (0,5), (1,0), (1,1), (1,3), (1,4), (1,5), (1,6), (2,0), (2,2), (2,3), (2,7), (3,1), (3,4), (3,5). Wave numbers of the (1,0), (2,0) and (3,1) bands are given in the tables 1-3. Some bands of the β -system ($B^2\Pi-x^2\Pi$) were also present between the much stronger γ -bands; the (1,5), (1,6), (2,3), (2,4) and (4,2) β -bands were analysed up to $K \simeq 35-40$, i.e. rather high rotational quantum numbers. When photographed in active nitrogen, which is especially favourable for exciting the β -bands, they are observable only up to $K \simeq 25-30$ (Jenkins, Barton and Mulliken 1927).

Averages of the upper-state combination differences of all bands with common initial states were evaluated. The rotational constants have been determined by the usual graphical method, which yields the B_v constants given in table 4. The data are not accurate enough to determine the variation of the D_v constants with v' ; $D' = 6.0 \cdot 10^{-6} \text{ cm}^{-1}$ was therefore used throughout.

The vibrational terms included in table 4 are the heights of the first rotational levels with $J = \frac{1}{2}$ of the different rotational term series above the $x^2\Pi_1$ ($v=0, J = \frac{1}{2}$) ground level, determined graphically using all bands and branches. The rotational constant and the height of the first rotational level of the $D^2\Sigma, v=0$ state, evaluated from the $v'=0, \epsilon$ -bands, are also given in table 4. As can be seen, the B_0' of the ϵ -bands is greater than each of the B' -values of the γ -bands;

* This paper was written by L. Gerö in 1943. In early 1945, during the period after the siege of Budapest, the first copy of the manuscript went astray, and figures 1 and 2 had to be reconstructed. In figure 2 we were able to mark only schematically the places where the accidental predissociation effect was observed.—J. G. VALATIN.

Table 1

$\Lambda^2\Sigma - \chi^2\Pi_1; 1,0 \text{ band}$					$\Lambda^2\Sigma - \chi^2\Pi_1; 1,0 \text{ band}$				
J	R_2	R_1	Q_1	P_1	J	R_2	Q_2	P_2	OP_{12}
1 $\frac{1}{2}$	46552.28	46544.61	46540.51		1 $\frac{1}{2}$	46439.04	46427.87	46419.77	46415.95
1 $\frac{3}{2}$	558.89	547.66	539.28	46535.57	2 $\frac{1}{2}$	446.70	431.12	419.08	411.21
2 $\frac{1}{2}$	566.48	551.09	538.73	531.18	2 $\frac{3}{2}$	454.34	434.81	419.08	407.16
3 $\frac{1}{2}$	574.70	554.97	538.73	527.39	3 $\frac{1}{2}$	462.26	439.04	419.08	403.31
4 $\frac{1}{2}$	583.36	559.61	539.28	524.27	4 $\frac{1}{2}$	470.97	443.83	419.77	400.31
5 $\frac{1}{2}$	592.20	564.85	541.10	521.58	5 $\frac{1}{2}$	480.04	448.97	421.06	397.60
6 $\frac{1}{2}$	602.34	570.91	543.13	519.40	6 $\frac{1}{2}$	490.00	454.34	422.91	395.42
7 $\frac{1}{2}$	612.57	577.11	545.50	517.84	7 $\frac{1}{2}$	500.19	460.71	425.13	393.87
8 $\frac{1}{2}$	623.42	584.21	548.57	517.08	8 $\frac{1}{2}$	510.46	467.19	427.87	392.45
9 $\frac{1}{2}$	634.88	591.52	552.28	517.08	9 $\frac{1}{2}$	521.58	474.36	431.12	391.87
10 $\frac{1}{2}$	646.85	599.84	556.44	517.08	10 $\frac{1}{2}$	533.10	482.01	434.81	391.87
11 $\frac{1}{2}$	659.67	608.51	561.45	517.84	11 $\frac{1}{2}$	545.50	490.00	439.04	391.87
12 $\frac{1}{2}$	673.24	617.92	566.48	519.40	12 $\frac{1}{2}$	557.81	498.91	443.83	392.45
13 $\frac{1}{2}$	686.40	627.87	572.56	521.58	13 $\frac{1}{2}$	570.91	508.00	448.97	393.87
14 $\frac{1}{2}$	700.99	638.34	579.19	524.27	14 $\frac{1}{2}$	584.21	517.84	454.88	395.81
15 $\frac{1}{2}$	716.11	649.35	586.44	527.39	15 $\frac{1}{2}$	598.62	528.02	461.23	398.25
16 $\frac{1}{2}$	731.31	660.87	594.04	531.18	16 $\frac{1}{2}$	613.55	538.73	467.92	401.22
17 $\frac{1}{2}$	747.65	673.24	602.34	535.57	17 $\frac{1}{2}$	628.35	549.83	475.22	404.58
18 $\frac{1}{2}$	764.47	686.07	611.05	540.51	18 $\frac{1}{2}$	643.97	561.45	483.06	408.54
19 $\frac{1}{2}$	781.62	699.27	620.53	546.02	19 $\frac{1}{2}$	659.67	573.99	491.26	413.00
20 $\frac{1}{2}$	799.31	713.23	630.66	552.28	20 $\frac{1}{2}$	676.68	586.44	500.19	418.00
21 $\frac{1}{2}$	817.79	727.71	641.19	558.89	21 $\frac{1}{2}$	694.01	599.84	509.55	422.91
22 $\frac{1}{2}$	836.76	743.06	652.35	566.48	22 $\frac{1}{2}$	711.66	613.55	519.40	429.21
23 $\frac{1}{2}$	856.39	758.50	664.16	573.99	23 $\frac{1}{2}$	729.87	627.87	529.68	435.50
24 $\frac{1}{2}$	876.47	774.71	676.68	582.34	24 $\frac{1}{2}$	748.40	642.82	540.51	442.90
25 $\frac{1}{2}$	897.11	791.50	689.26	591.52	25 $\frac{1}{2}$	767.64	658.01	552.28	450.11
26 $\frac{1}{2}$	918.24	808.76	702.67	600.75	26 $\frac{1}{2}$	787.47	674.13	564.14	458.40
27 $\frac{1}{2}$	939.74	826.74	716.55	611.05	27 $\frac{1}{2}$	807.64	690.30	576.63	467.19
28 $\frac{1}{2}$	962.12	845.01	731.31	621.64	28 $\frac{1}{2}$	828.36	707.05	589.46	
29 $\frac{1}{2}$	985.18	864.09	746.32	633.07	29 $\frac{1}{2}$	849.60	724.54	602.92	485.53
30 $\frac{1}{2}$	47008.58	883.59	761.97	644.67	30 $\frac{1}{2}$	871.39	742.36	617.16	495.58
31 $\frac{1}{2}$	032.46	903.87	778.21	657.32	31 $\frac{1}{2}$	893.72	760.84	631.36	506.40
32 $\frac{1}{2}$	057.01	924.56	795.11	670.21	32 $\frac{1}{2}$	916.57	779.80	646.85	517.84
33 $\frac{1}{2}$	082.05	945.58	812.44	683.68	33 $\frac{1}{2}$	939.74	799.31	662.17	529.68
34 $\frac{1}{2}$	107.87	967.26	830.40	697.63	34 $\frac{1}{2}$	963.51	819.35	678.60	
35 $\frac{1}{2}$	133.96	989.66	848.90	712.07	35 $\frac{1}{2}$	987.72	839.77	695.08	554.97
36 $\frac{1}{2}$		47012.55	867.92	727.71	36 $\frac{1}{2}$	47012.55	860.87	712.07	567.91
37 $\frac{1}{2}$		035.93	887.52	743.06	37 $\frac{1}{2}$	038.08	882.31	729.87	581.65
38 $\frac{1}{2}$		059.84	907.67	759.44	38 $\frac{1}{2}$	063.94	904.48	748.40	595.98
39 $\frac{1}{2}$		084.53	928.37	776.39	39 $\frac{1}{2}$	090.34	927.25	767.01	611.05
40 $\frac{1}{2}$		109.46	949.50	793.88	40 $\frac{1}{2}$	117.33	950.12	786.35	626.59
41 $\frac{1}{2}$		135.18	971.57	811.65	41 $\frac{1}{2}$	144.74	973.83	806.15	642.82
42 $\frac{1}{2}$		161.27	993.57	830.40	42 $\frac{1}{2}$	172.70	998.03	826.74	659.67
43 $\frac{1}{2}$		187.93	47016.49	849.60	43 $\frac{1}{2}$	201.27	47022.48	847.56	676.68
44 $\frac{1}{2}$		215.34	040.09	868.99	44 $\frac{1}{2}$	230.20	047.93	868.99	694.01
45 $\frac{1}{2}$		243.07	063.94	889.35	45 $\frac{1}{2}$	259.73	073.73	890.82	
46 $\frac{1}{2}$		271.35	088.47	910.14	46 $\frac{1}{2}$	290.00	099.99	913.28	
47 $\frac{1}{2}$		300.20	113.59	931.59	47 $\frac{1}{2}$	320.37	126.83	936.39	
48 $\frac{1}{2}$		329.47	139.39	953.10	48 $\frac{1}{2}$	351.48	154.21	959.80	
49 $\frac{1}{2}$		359.64	165.49	975.51	49 $\frac{1}{2}$	383.12	182.01	983.81	
50 $\frac{1}{2}$		390.10	192.23	998.03	50 $\frac{1}{2}$	415.32	210.39	47008.58	
51 $\frac{1}{2}$		421.32	219.58	47022.48	51 $\frac{1}{2}$	447.99	239.37	033.79	
52 $\frac{1}{2}$		452.83	247.36	046.07	52 $\frac{1}{2}$	481.14	268.80	059.84	
53 $\frac{1}{2}$		485.00	275.69	070.78	53 $\frac{1}{2}$	514.78	298.66	085.79	
54 $\frac{1}{2}$		517.63	304.65	095.97	54 $\frac{1}{2}$	549.29	329.47	112.55	
55 $\frac{1}{2}$		550.85	334.12	121.77	55 $\frac{1}{2}$	583.78	360.36	139.39	
56 $\frac{1}{2}$		584.48	364.17	148.08	56 $\frac{1}{2}$	618.93	391.86	167.76	
57 $\frac{1}{2}$		618.93	394.78	174.82	57 $\frac{1}{2}$	654.90	424.11	196.04	
58 $\frac{1}{2}$		653.45	425.89	202.01	58 $\frac{1}{2}$	691.20	456.71	224.97	
59 $\frac{1}{2}$		689.02	457.43	230.20	59 $\frac{1}{2}$	728.20	489.62	254.37	
60 $\frac{1}{2}$		724.90	489.62	258.46	60 $\frac{1}{2}$	765.44	523.59	284.65	
61 $\frac{1}{2}$		760.94	522.46	287.60	61 $\frac{1}{2}$	803.37	557.82	314.93	
62 $\frac{1}{2}$		798.25	555.62	317.19	62 $\frac{1}{2}$	841.30	592.66	346.05	
63 $\frac{1}{2}$		835.75	589.44	347.49	63 $\frac{1}{2}$	880.47	627.70	377.66	
64 $\frac{1}{2}$		873.78	623.69	378.74	64 $\frac{1}{2}$	919.90	663.65	409.50	
65 $\frac{1}{2}$		912.34	658.95	409.50	65 $\frac{1}{2}$		669.62	442.39	
66 $\frac{1}{2}$		951.20	694.07	440.88	66 $\frac{1}{2}$			475.55	
67 $\frac{1}{2}$			729.98	473.14	67 $\frac{1}{2}$				
68 $\frac{1}{2}$				505.60					

Table 2

$A^2\Sigma - x^2\Pi_1; 2,0$ band					$A^2\Sigma - x^2\Pi_1; 2,0$ band				
J	$S_{R_{21}}$	R_1	Q_1	P_1	J	R_2	Q_2	P_2	oP_{11}
$\frac{1}{2}$	48861.73	48854.14	48850.11		$1\frac{1}{2}$		48737.33	48729.52	
$1\frac{1}{2}$	868.28	856.62	849.08	48844.83	$2\frac{1}{2}$		739.90	728.59	
$2\frac{1}{2}$	875.42	860.16	848.29	839.85	$3\frac{1}{2}$	48763.05	743.46	727.90	48716.22
$3\frac{1}{2}$	883.30	863.88	848.29	836.10	$4\frac{1}{2}$	771.40	747.43	727.90	712.35
$4\frac{1}{2}$	891.89	868.28	849.08	832.78	$5\frac{1}{2}$	779.30	752.00	728.59	709.08
$5\frac{1}{2}$	900.71	873.82	850.11	830.33	$6\frac{1}{2}$	788.39	756.89	729.52	706.28
$6\frac{1}{2}$	910.23	879.70	851.92	828.06	$7\frac{1}{2}$	797.35	762.27	731.08	703.67
$7\frac{1}{2}$	920.38	885.05	854.14	826.39	$8\frac{1}{2}$	807.06	768.12	733.01	701.53
$8\frac{1}{2}$	930.65	891.89	856.62	825.08	$9\frac{1}{2}$	817.50	774.52	735.31	700.75
$9\frac{1}{2}$	941.81	898.97	860.16	824.03	$10\frac{1}{2}$	828.01	781.23	738.33	699.16
$10\frac{1}{2}$	953.43	907.00	863.88	824.03	$11\frac{1}{2}$	839.23	788.39	741.68	698.87
$11\frac{1}{2}$	965.54	915.21	868.28	825.08	$12\frac{1}{2}$	850.51	796.20	745.43	698.87
$12\frac{1}{2}$	978.40	923.83	873.08	826.39	$13\frac{1}{2}$	863.09	804.42	749.76	699.16
$13\frac{1}{2}$	991.93	933.09	878.77	828.06	$14\frac{1}{2}$	875.42	812.97	754.42	700.75
$14\frac{1}{2}$	49005.38	943.06	884.67	830.33	$15\frac{1}{2}$	888.09	822.06	759.70	701.53
$15\frac{1}{2}$	019.58	953.43	891.10	832.78	$16\frac{1}{2}$	901.60	831.54	765.33	702.93
$16\frac{1}{2}$	034.56	964.58	898.08	836.10	$17\frac{1}{2}$	915.21	841.61	771.40	705.07
$17\frac{1}{2}$	049.88	975.97	905.85	839.85	$18\frac{1}{2}$	929.91	851.92	778.21	708.06
$18\frac{1}{2}$	065.75	988.14	914.09	844.14	$19\frac{1}{2}$	944.66	863.09	785.23	711.33
$19\frac{1}{2}$	082.25	49000.79	922.85	849.08	$20\frac{1}{2}$	960.07	874.54	792.70	714.91
$20\frac{1}{2}$	099.36	013.84	932.04	854.14	$21\frac{1}{2}$	975.97	886.34	800.75	719.19
$21\frac{1}{2}$	117.04	027.49	941.81	860.16	$22\frac{1}{2}$	991.93	898.97	809.21	723.63
$22\frac{1}{2}$		041.99	952.24	866.77	$23\frac{1}{2}$	49008.67	911.76	818.34	728.59
$23\frac{1}{2}$		056.48	963.10	873.82	$24\frac{1}{2}$	025.73	924.98	828.06	734.69
$24\frac{1}{2}$		071.86	974.25	881.32	$25\frac{1}{2}$	043.63	938.87	837.65	
$25\frac{1}{2}$		087.55	986.20	889.28	$26\frac{1}{2}$	061.63	953.43	848.29	
$26\frac{1}{2}$		103.55	998.85	898.08	$27\frac{1}{2}$	080.39	967.89	859.25	
$27\frac{1}{2}$		120.71	49011.58	907.00	$28\frac{1}{2}$	099.36	983.15	870.50	
$28\frac{1}{2}$		137.99	025.28	916.86	$29\frac{1}{2}$	119.09	998.85	882.32	
$29\frac{1}{2}$		155.56	039.30	926.87	$30\frac{1}{2}$	138.98	49015.02	894.78	
$30\frac{1}{2}$		174.31	053.80	937.67	$31\frac{1}{2}$	159.35	031.79	907.58	
$31\frac{1}{2}$		193.09	068.87	949.05	$32\frac{1}{2}$	180.12	048.94	920.93	
$32\frac{1}{2}$		212.61	084.56	960.75	$33\frac{1}{2}$	201.76	066.56	934.82	
$33\frac{1}{2}$		232.31	100.39	972.86	$34\frac{1}{2}$	223.75	084.56	949.05	
$34\frac{1}{2}$		252.63	117.04	986.20	$35\frac{1}{2}$	246.30	103.55	963.91	
$35\frac{1}{2}$		273.65	134.15	998.85	$36\frac{1}{2}$	269.72	122.54	979.21	
$36\frac{1}{2}$		295.18	152.12	49012.71	$37\frac{1}{2}$		142.24	995.01	
$37\frac{1}{2}$		317.13	170.20	027.49	$38\frac{1}{2}$	316.40		49011.58	
$38\frac{1}{2}$		339.13	188.78	041.99	$39\frac{1}{2}$	340.45	182.71		
$39\frac{1}{2}$		362.66	207.57	057.66	$40\frac{1}{2}$	365.27	203.96	045.36	
$40\frac{1}{2}$		385.88	227.80	073.26	$41\frac{1}{2}$	390.54	225.37	063.07	
$41\frac{1}{2}$		410.36	248.00	090.13	$42\frac{1}{2}$	416.51	247.51	081.21	
$42\frac{1}{2}$		434.71	268.96	106.98	$43\frac{1}{2}$	442.58	269.72	099.36	
$43\frac{1}{2}$		459.95	290.00	124.52	$44\frac{1}{2}$	469.74	292.77	119.09	
$44\frac{1}{2}$		485.14	311.69	142.24	$45\frac{1}{2}$	497.27	316.30	138.98	
$45\frac{1}{2}$		511.26	334.12	161.14	$46\frac{1}{2}$	525.91	340.45	159.35	
$46\frac{1}{2}$		537.63	356.74	180.12	$47\frac{1}{2}$	553.52	365.27	180.12	
$47\frac{1}{2}$		564.89	380.17	199.82	$48\frac{1}{2}$	582.74	390.54	201.22	
$48\frac{1}{2}$		592.94	404.01	219.68	$49\frac{1}{2}$	612.31	416.51	222.93	
$49\frac{1}{2}$		621.05	428.19	240.62	$50\frac{1}{2}$	642.36	442.58	244.94	
$50\frac{1}{2}$		649.14	452.98	261.03	$51\frac{1}{2}$	673.32	469.74	267.63	
$51\frac{1}{2}$		678.99	478.38	282.42	$52\frac{1}{2}$		497.27		
$52\frac{1}{2}$			504.47	303.66					
$53\frac{1}{2}$				325.65					

this is a strong argument against the hypothesis that the γ and ϵ bands belong to the same band system. It must also be determined whether there are homogeneous perturbations in one or in another of these bands which deform the rotational constants. From figure 1 it is seen that the $B'-B''$ curves (Gerö 1935 a) for the bands γ (0,2), (1,0), (2,0) and (3,1) and ϵ (0,3) are nearly horizontal straight lines which means that no considerable perturbations are present in these bands.

Predissociation phenomena were found in the upper state of the γ -bands. The rotational term series show sudden intensity falls at $v=0$, $K=74$; $v=1$,

Table 3

$\Lambda^2\Sigma - x^2\Pi_1; 3,1 \text{ band}$				$\Lambda^2\Sigma - x^2\Pi_1; 3,1 \text{ band}$			
J	R_1	Q_1	P_1		R_2	Q_2	P_2
$\frac{1}{2}$	49254.40			$\frac{1}{2}$			
$1\frac{1}{2}$	257.02	49249.43		$1\frac{1}{2}$		49137.99	49130.29
$2\frac{1}{2}$	260.52	249.43		$2\frac{1}{2}$		140.48	129.01
$3\frac{1}{2}$	264.21	249.43		$3\frac{1}{2}$		144.22	129.01
$4\frac{1}{2}$	268.96	249.43		$4\frac{1}{2}$		147.99	129.01
$5\frac{1}{2}$	273.65	250.23	49231.53	$5\frac{1}{2}$		152.12	129.01
$6\frac{1}{2}$	278.90	252.12	228.91	$6\frac{1}{2}$		157.49	130.29
$7\frac{1}{2}$	284.87	254.40	227.80	$7\frac{1}{2}$		162.95	131.98
$8\frac{1}{2}$	291.43	257.02	226.05	$8\frac{1}{2}$		168.61	133.09
$9\frac{1}{2}$	298.46	260.52	225.37	$9\frac{1}{2}$		174.82	135.99
$10\frac{1}{2}$	306.34	264.21	225.37	$10\frac{1}{2}$		181.60	138.98
$11\frac{1}{2}$	314.75	268.96	226.05	$11\frac{1}{2}$		188.78	142.24
$12\frac{1}{2}$	323.60	273.65	227.80	$12\frac{1}{2}$		196.36	146.23
$13\frac{1}{2}$	332.80	278.90	228.91	$13\frac{1}{2}$		203.96	150.41
$14\frac{1}{2}$	342.66	284.87	231.53	$14\frac{1}{2}$		212.61	155.56
$15\frac{1}{2}$	353.04	291.43	233.71	$15\frac{1}{2}$	49287.60	222.05	160.44
$16\frac{1}{2}$	363.74	298.46	236.91	$16\frac{1}{2}$	300.74	231.53	166.09
$17\frac{1}{2}$	375.34	305.95	240.62	$17\frac{1}{2}$	314.75	241.45	172.23
$18\frac{1}{2}$	387.19	314.14	244.94	$18\frac{1}{2}$	328.89	252.12	179.01
$19\frac{1}{2}$	399.76	322.73	249.43	$19\frac{1}{2}$	343.66	262.61	185.90
$20\frac{1}{2}$	412.99	331.93	255.13	$20\frac{1}{2}$	358.67	274.16	193.09
$21\frac{1}{2}$	426.45	341.49	261.03	$21\frac{1}{2}$	373.98	285.92	201.22
$22\frac{1}{2}$	440.77	352.14	267.63	$22\frac{1}{2}$	390.54	298.46	209.70
$23\frac{1}{2}$	454.86	362.66	274.16	$23\frac{1}{2}$	407.11	310.98	218.81
$24\frac{1}{2}$	469.74	373.98	281.98	$24\frac{1}{2}$	424.50	324.29	228.91
$25\frac{1}{2}$	485.14	385.88	290.00	$25\frac{1}{2}$	441.68	338.09	238.15
$26\frac{1}{2}$	501.75	397.91	298.46	$26\frac{1}{2}$	459.95	352.14	248.00
$27\frac{1}{2}$	518.74	410.99	307.58	$27\frac{1}{2}$	478.38	366.88	259.38
$28\frac{1}{2}$	535.99	424.50	317.13	$28\frac{1}{2}$	497.27	382.15	270.62
$29\frac{1}{2}$	553.52	438.28	327.50	$29\frac{1}{2}$	516.56	397.91	282.42
$30\frac{1}{2}$	572.12	452.98	338.09	$30\frac{1}{2}$		413.77	295.18
$31\frac{1}{2}$	590.71	467.22	348.97	$31\frac{1}{2}$		430.31	307.58
$32\frac{1}{2}$	609.77	483.03	360.67	$32\frac{1}{2}$		447.21	320.73
$33\frac{1}{2}$	629.99	498.84	373.03	$33\frac{1}{2}$		464.62	334.12
$34\frac{1}{2}$	650.01	515.90	385.88	$34\frac{1}{2}$		483.03	348.97
$35\frac{1}{2}$	670.63	532.78	398.96	$35\frac{1}{2}$		501.75	363.74
$36\frac{1}{2}$	692.03	550.25	412.99	$36\frac{1}{2}$			380.17
$37\frac{1}{2}$		568.31	427.50				
$38\frac{1}{2}$			441.68				

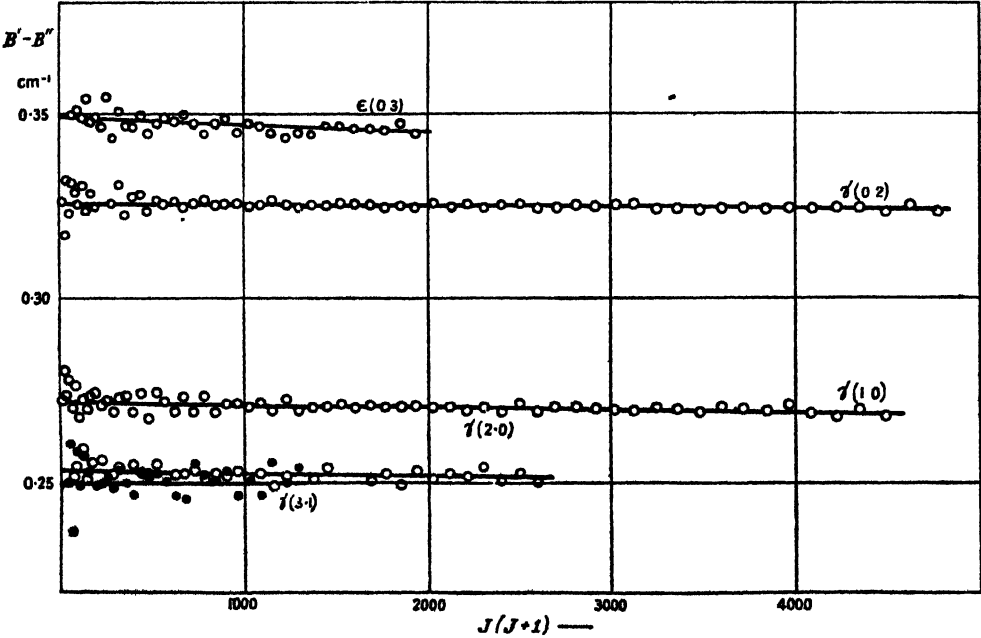


Table 4

		Vibrational terms			B_v
$v=0$	44199.2	2341.4			1.9870
1	46540.6	2309.9	31.5		1.9688
$\Lambda^2\Sigma$	2	2276.7	33.2	1.7	1.9498
3	51127.2				1.9290
$\nu^2\Sigma$ $v=0$	53291.9				1.9917

$K=64$; $v=2$, $K=52$; $v=3$, $K=38$; joining these points in figure 2, a limiting curve of predissociation can be drawn, which intersects the ordinate axis at a value of about $53\,800\text{ cm}^{-1}$. An $N+O$ atomic term combination must lie at this energy, which gives rise to the predissociation phenomena observed.

The upper-state rotational term series of the β -bands intersect in many places the much steeper upper term series of the γ -bands. In the neighbourhood of the intersections perturbation phenomena should be discovered. Tables 5 and 6 contain parts of the corresponding branches of the $(4,2)\beta$ – $(2,2)\gamma$ and $(2,4)\beta$ – $(1,4)\gamma$ bands respectively, which belong to common lower states. At the intersection points of the upper states the differences of corresponding wave numbers change their signs; these places in the tables are marked with arrows.

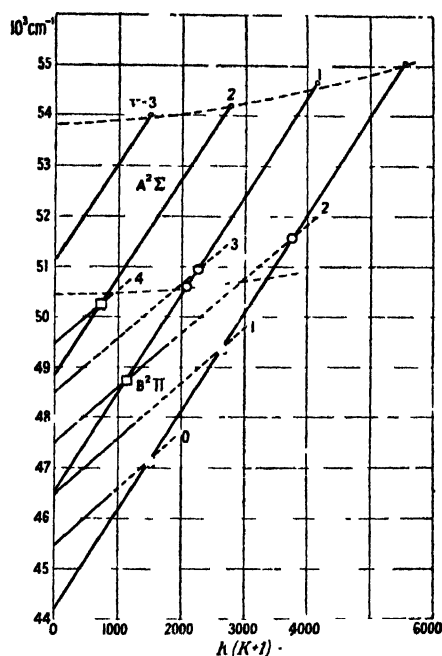


Figure 2.

Table 5

J	$\beta(4,2)$ R_1	$\gamma(2,2)$ R_1	$\beta(4,2)$ P_1	$\gamma(2,2)$ P_1	$\beta(4,2)$ R_2	$\gamma(2,2)$ R_2
$21\frac{1}{2}$	45519.76	45320.10	45428.05	45153.05	45440.58	45269.54
$22\frac{1}{2}$	495.79	335.97	399.50	161.19	416.22	287.42
$23\frac{1}{2}$	469.47	352.28	369.32	169.72	390.21	305.87
$24\frac{1}{2}$	442.75	369.32	338.31	178.49	364.38	324.56
$25\frac{1}{2}$	414.45	386.66	305.87	188.60	336.77	344.22
$26\frac{1}{2}$	384.74	404.90	272.49	199.12	307.58	364.38
$27\frac{1}{2}$	354.48	423.63	237.61	210.06	277.62	384.74
$28\frac{1}{2}$	323.17	442.75	202.43	221.63	246.49	405.79
$29\frac{1}{2}$	290.44	462.84	165.15	233.76		
$30\frac{1}{2}$		483.19	127.00	246.49		

No irregularity of any kind could be detected in the wave numbers or intensities in either of the branches.

The $R_1(32\frac{1}{2})$ and $P_1(34\frac{1}{2})$ lines of the $(2,4)\beta$ and $(1,4)\gamma$ bands respectively have the same wave numbers; thus the upper rotational terms $F_1(33\frac{1}{2})$ of both

Table 6

J	$\gamma(1,4)$ $S_{R_{11}}$	$\beta(2,4)$ R_1	$\gamma(1,4)$ R_1	$\beta(2,4)$ P_1	$\gamma(1,4)$ F_1	$\beta(2,4)$ R_2	$\gamma(1,4)$ R_2	$\beta(2,4)$ P_2	$\gamma(1,4)$ F_2
26 $\frac{1}{2}$	39632.27	39841.66	39522.52	39726.52	39314.99	39763.22	39484.56	39644.29	39269.28
27 $\frac{1}{2}$	657.40	814.48	544.28	695.15	329.11	736.06	508.24	612.92	285.03
28 $\frac{1}{2}$	683.82	786.20	566.69	662.55	343.21	707.74	532.53	580.19	301.59
29 $\frac{1}{2}$	710.85	756.76	589.74	628.92	359.05	678.48	557.58	546.47	318.93
30 $\frac{1}{2}$	738.54	726.52	613.60	594.53	374.97	647.97	583.19	511.90	336.78
31 $\frac{1}{2}$	767.05	695.15	638.00	558.75	391.65	616.45	609.48	476.03	355.35
32 $\frac{1}{2}$	796.00	663.26	663.26	522.07	409.13	583.84	636.37	439.48	374.57
33 $\frac{1}{2}$	826.04	629.96	689.33	484.56	427.27	550.11	663.82	401.41	394.40
34 $\frac{1}{2}$	856.61	595.71	715.91	445.99	445.99	515.35	692.15	362.24	415.02
35 $\frac{1}{2}$	887.14	559.58	743.07	405.96	465.37	479.57	721.14	322.52	436.23

bands have almost equal energies. As neither line broadening, nor anomalous intensity are observable (the intensity of the common lines being approximately the sum of the intensities of the two coinciding lines), the interaction of the two electronic states $B^3\Pi$ and $A^3\Sigma$ must be extraordinarily small.

The points at which the expected perturbations could not be detected are

marked with squares in figure 2, full lines show where both intersecting term series were observed, and broken lines, where only the upper series of the γ -bands were observed. There are, however, three points where intensity anomalies were observed in the branches of the γ -bands; these points are marked with circles. Two circles lying close together in a term series mean that as a result of intersection with Π -states, the F_1 and F_2 perturbations of the $A^2\Sigma$ term series do not take place at the same J -values. These points also lie near the γ - β upper-term intersections, but at higher energies than those intersections at which no effect could be detected. Between the circles and squares a limiting curve (similar to the limiting curve of predissociation at $53\,800\text{ cm}^{-1}$) can be drawn. Above this limiting curve, intensity diminutions are present, but no shifts of rotational level; below the curve no such effects are observed. As the (4,2) band of the β system comes to an end with a sudden intensity drop just at this limiting curve and no β -bands with $v' > 4$ were found, the limit at $50\,500\text{ cm}^{-1}$ is a real predissociation limit. This limit corresponds to a continuum, which is in strong interaction with the $B^2\Pi$ state. At the intersection points above the limiting curve, the eigenfunctions of the $B^2\Pi$ state are mixed with the eigenfunctions of the $A^2\Sigma$ state and this mediates the intensity-weakening for the γ -band lines. Such a combination of predissociation and perturbation effects was called by Coster, Brons and van der Ziel (1933) "accidental predissociation" and was treated theoretically by Ittmann (1934) in the case of the Second Positive bands of the N_2 molecule; but neither the perturbing states, nor the cause of predissociation were known in the N_2 spectrum, while in NO both are well established. (See also Gerö 1935 b, and Gerö and Schmid 1940.)

The energy difference of the two predissociation limits found is about 3300 cm^{-1} . Now in the energy scheme of $N + O$ atomic term combinations, only one difference

comes near to this value, namely the difference $N(^3D) + O(^3P) - N(^4S) + O(^1D) = 3355\text{ cm}^{-1}$; all other differences are much larger. Relating these two combinations to the limiting curves, the term scheme of figure 3 is obtained. According to this scheme, the dissociation energy of the NO molecule is $34\,600\text{ cm}^{-1} = 4.29\text{ eV}$. The $x^2\Pi$ ground state dissociates *not* into the $N(^4S) + O(^3P)$ ground atomic term combination, but into the $N(^3D) + O(^3P)$ combination, with a dissociation energy of $53\,800\text{ cm}^{-1}$. (The extrapolated value, which in most cases is 10–25 % greater than the real one, is $\omega_e^2/(4\omega_e x_e) = 62\,500\text{ cm}^{-1}$, and, therefore, a dissociation energy of $34\,600\text{ cm}^{-1}$ for the ground state would be too small.)

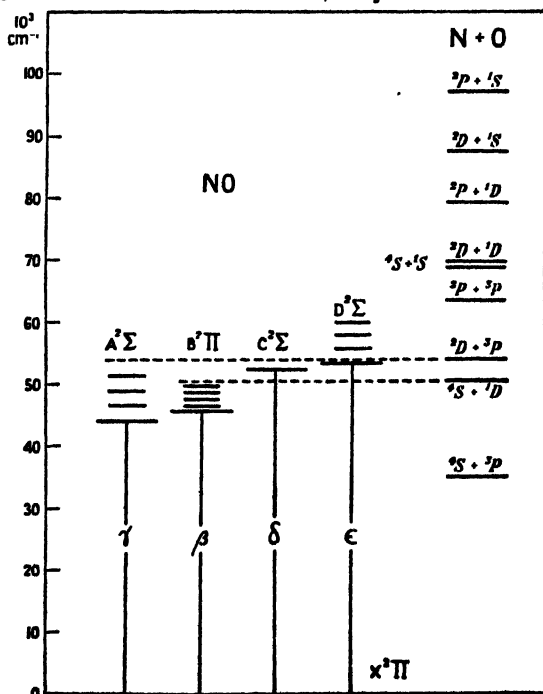


Figure 3.

The above dissociation scheme of NO is strongly supported by the photochemical experiments of Flory and Johnston (1935). They irradiated NO with the 1832 Å line of mercury and found considerable photochemical decomposition. The energy of this mercury line, $54\,567\text{ cm}^{-1}$, is slightly greater than the dissociation limit of the NO ground state at $53\,800\text{ cm}^{-1}$; hence the efficiency of this wavelength.

REFERENCES

- COSTER, D., BRONS, F., and VAN DER ZIEL, A., 1933, *Z. Phys.*, **84**, 304.
 FLORY, P. J., and JOHNSTON, H. L., 1935, *J. Amer. Chem. Soc.*, **57**, 2641.
 GERÖ, L., 1935 a, *Z. Phys.*, **93**, 669 ; 1935 b, *Ibid.*, **96**, 669.
 GERÖ, L., and SCHMID R., 1940, *Z. Phys.*, **116**, 246.
 GERÖ, L., SCHMID, R., and VON SZILY, F. K., 1944, *Physica*, **11**, 144.
 ITTMANN, G. P., 1934, *Naturwissenschaften*, **22**, 118.
 JENKINS, F. A., BARTON, H. A., and MULLIKEN, R. S., 1927, *Phys. Rev.*, **30**, 150, 175.

Collision Broadening of the Inversion Spectrum of Ammonia: III. The Collision Cross-sections for Self-broadening and for Mixtures with Non-polar Gases

BY B. BLEANEY AND R. P. PENROSE

Clarendon Laboratory, Oxford

ABSTRACT. The widths of the lines of the inversion spectrum of pure ammonia near 1 cm^{-1} previously determined experimentally (Bleaney and Penrose 1947) are shown to vary from line to line in a manner consistent with the assumption of a simple dipole-dipole interaction for the collision mechanism. Measurements of the line breadth constant for the line (J,K) - (3,3) in mixtures of ammonia with six different non-polar gases are described, and from them the collision diameters are calculated. They are found to be approximately the same as the diameters obtained from the kinetic theory, whereas in pure ammonia the diameters are two to four times greater.

A survey of microwave spectra indicates that the collision diameter is only significantly greater than the kinetic theory diameter for encounters between permanent electric dipoles.

§1. INTRODUCTION

THE analysis of the ammonia spectrum has been described in an earlier paper (Bleaney and Penrose 1947, hereafter called "A"). Of twenty-nine lines which were identified, seventeen were so well resolved at a pressure of 0.5 mm. Hg that they could be examined in detail. By measuring the absorption coefficient at a number of frequencies the shape of each line could be determined and the line breadth constant calculated. The results of these measurements were given in table 4 of paper A; for convenience, this table is repeated here (table 1), with the order rearranged so that the lines are given in order of increasing breadth.

It will be seen that the values of the line breadth constant at the pressure of 0.5 mm. Hg vary from $1.8 \times 10^{-4}\text{ cm}^{-1}$ to $4.7 \times 10^{-4}\text{ cm}^{-1}$. From them the appropriate collision frequencies can be calculated since the collision frequency f is associated with a line breadth constant $\Delta\bar{\nu} = f/2\pi c$. This collision frequency,

which is the number of times per second that the absorption of radiation by one molecule is interrupted by the approach of another molecule, varies from 3.4×10^7 to 8.9×10^7 per second. The collision frequency at 0.5 mm. pressure deduced from viscosity measurements by means of the kinetic theory is only 4.0×10^6 per second, which is considerably smaller. It follows that the interruption of absorption of centimetre-wave radiation occurs when another molecule approaches to a distance which is considerably greater than that required for an appreciable transfer of momentum from one molecule to the other. The relative distances can be found by calculating the collision diameters (σ) of the ammonia molecule from the formula $f = \sqrt{2}v\pi\sigma^2N$, where N is the number of molecules per cm^3 and v is the average velocity of the molecules. The values of σ range from 8.7 Å. for the narrowest lines to 14.2 Å. for the broadest, whereas the kinetic theory diameter is only 4.4 Å.

Table 1

Line (J, K)	$\Delta\bar{\nu}$ at $p=0.5$ mm. Hg $\times 10^{-4} \text{ cm}^{-1}$	γ_0 (db/km.)	I_{exp} db/mol	I_{theor} db/mol
5,1	1.8	7	3.3	3.5
3,1	2.4	29	14.5	14
2,1	2.6	51	25	27
5,2	2.6	20	13	15
11,9	2.9	8.5	5.7	6.4
6,3	3.1	44	35.5	34
3,2	3.2	100	62.5	59
5,3	3.3	96	71	73.5
6,4	3.6	42	35	34.5
7,6	3.9	132	100	95.5
7,5	4.1	28	27	27
8,7	4.1	44	34	32.5
10,9	4.2	35	25.5	22
3,3	4.5	260	289	289
4,4	4.5	190	149	151
5,5	4.7	186	148	137
6,6	4.7	276	198	210

$\Delta\bar{\nu}$ is measured value of the line breadth constant (cm^{-1}).

γ_0 is measured value of absorption at centre of line (db/km.).

I is integrated intensity = $\int_0^\infty \frac{\gamma}{\bar{\nu}^2} d\bar{\nu}$ reduced to a length of path containing one mol of gas per unit cross-section.

I_{exp} is calculated from the relation $I_{\text{exp}} = \bar{\nu}_0^2 \pi \gamma_0 \Delta\bar{\nu} (10^{-8} V_0)$, where V_0 = molar volume at a pressure of 0.5 mm. Hg.

I_{theor} is calculated from the relation $I_{\text{theor}} = \frac{4\pi^3 N_{JK} V_0 |\mu_{JK}|^2}{3kT}$

N_{JK} = number of ammonia molecules per cm^3 in rotational level (J, K).

$|\mu_{JK}|^2$ = square of dipole moment associated with the line (J, K).

These large values for the collision cross-section may be attributed to the effect of the large electric field of the ammonia dipole, which will produce a considerable perturbation of the energy levels of an approaching molecule at distance of the order of 10^{-7} cm. This assumption is supported by consideration of the large variation in the widths of the different lines. This is at first sight surprising, since each line arises from an exactly similar transition between the levels of the inversion doublet. The significant difference arises, however, from the fact

that each line is due to molecules in a particular rotational energy level, and the line widths vary with the state of rotation of the molecule in the same way as does the Stark splitting due to an external electric field. In the following section of this paper the variation of line width is discussed, and it is shown that it can be explained by the simple hypothesis that a collision occurs when the electrical forces between the molecules due to their strong dipoles cause the energy of interaction to reach a certain value. The magnitude of the interaction energy required to produce a collision is evaluated from the observed line widths by averaging over the various settings of the dipoles with respect to the line of centres at a collision.

§2. THE COLLISION CROSS-SECTION FOR SELF-BROADENING

The magnitude of the splitting or perturbation of the energy levels of a molecule possessing a permanent electrical dipole moment when subjected to a constant electric field is determined by the steady component of the dipole moment. In the general state of rotation of a symmetrical top molecule, the dipole moment will be precessing about the axis of total angular momentum of the molecule, and the steady component is determined by the projection of the moment on that axis. The frequency of precession is of the same order as the rotational frequency, and one may therefore expect that the perturbation due to a transient electric field is still determined by the steady component of the dipole moment if the duration of the transient is long compared to the period of rotation. Since the experimental collision diameter is about 10^{-7} cm. for ammonia, the duration of a collision is of the order of 10^{-11} sec., while the rotational periods are less than 10^{-12} sec. Thus it is reasonable to assume that the interaction between the electric field of one molecule and the dipole moment of an absorbing molecule is determined by the steady component of the latter. In ammonia this varies as $K/(J^2 + J)^{1/2}$, where J, K are the rotational quantum numbers of a symmetrical top. The interaction should therefore be largest when $K=J$, and smallest when K is small compared with J . Inspection of table 1 shows that the line widths increase steadily as K/J approaches unity.

The variation of the line breadth constant with the rotational state of the molecule may be given a quantitative aspect by assuming that a collision occurs when two molecules approach to a distance such that the energy of interaction between them reaches a certain value W . For the molecules in a particular rotational state defined by the quantum numbers (J, K) , the average distance at which a collision takes place will then be determined by the relation $W \propto \bar{\mu}/r^3$. The variation in the line breadth constants depends only on the variation in the collision cross-section. Since the latter is proportional to r^2 , the line breadths should vary, on this hypothesis, as $(\bar{\mu}/W)^{2/3}$. Hence, assuming that W is independent of the rotational state of the molecule, $\Delta\bar{\nu}$ should vary as $(K^2/(J^2 + J))^{1/2}$ for the lines of this inversion spectrum. In figure 1 the widths

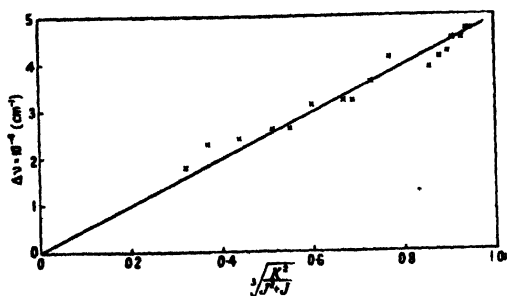


Figure 1. Widths of lines in ammonia inversion spectrum near 1 cm^{-1} .

($\Delta\bar{\nu}$) of the seventeen lines measured at 0.5 mm. Hg pressure are plotted against $(K^2/(J^2+J))^{\frac{1}{2}}$. Within the experimental error (about 5% in $\Delta\bar{\nu}$) the points fall on a straight line, thus confirming that the interruption of the absorption is due to dipole-dipole interaction.

The success of this simple theory encourages a more exact calculation of the magnitude of the interaction energy between two ammonia molecules required to produce a collision. This requires an averaging process in order to take account of the variation of the interaction energy between two dipoles with the angles at which they lie with respect to the line through their centres. A simple method of obtaining the average cross-section for collision is given in the Appendix to this paper. It is found that the average cross-sectional area for collisions of this type is

$$S_{11} = 2.45(\mu_1^2 \bar{\mu}_2^2 / W^2)^{\frac{1}{2}}, \quad \dots \dots (1)$$

where $\mu_1^2 = \mu^2 \{K^2/(J^2+J)\}$ is the squared dipole moment of an ammonia molecule in the rotational state (J, K) , $\bar{\mu}_2^2$ = the mean square of the dipole moment of all the ammonia molecules, and W = interaction energy required to produce a collision.

Hence the line breadth constant $\Delta\bar{\nu}$ is

$$\Delta\bar{\nu} = \frac{2NS_{11}}{\pi c} \sqrt{\left(\frac{RT}{M}\right)}, \quad \dots \dots (2)$$

where N is number of molecules per cm^3 , M is molecular weight, T absolute temperature, and c and R have their usual meanings.

At room temperature, computation of the average value of $\{K^2/(J^2+J)\}$ over all rotational states of ammonia gives $\bar{\mu}_2^2 = 0.40 \mu^2$. Inserting the expression for S_{11} , and taking $\mu = 1.44 \times 10^{-18}$ E.S.U., we obtain

$$\Delta\bar{\nu} = 4.5 \times 10^{-14} p_{\text{mm}} W^{-\frac{1}{2}} (K^2/(J^2+J))^{\frac{1}{2}} \quad (\text{cm}^{-1}), \quad \dots \dots (3a)$$

where p_{mm} is the pressure of ammonia in mm. Hg.

The experimental value of $\Delta\bar{\nu}$ at a pressure of 0.5 mm. Hg, obtained from figure 1, is

$$\Delta\bar{\nu} = 0.50 \times 10^{-3} (K^2/(J^2+J))^{\frac{1}{2}} \quad (\text{cm}^{-1}). \quad \dots \dots (3b)$$

Hence the interaction energy W is 3.0×10^{-16} ergs, or $W/hc = 1.5 \text{ cm}^{-1}$, which is approximately equal to twice the separation of the two energy levels of the inversion doublet, between which transitions are taking place. It is thus hardly surprising that the absorption or emission of radiation is interrupted.

The mechanism of this collision process is rather different from that assumed in the simple theory of collision broadening of spectral lines in the optical region. For such lines the periodic time (10^{-15} sec.) of the radiation is much shorter than the duration of a collision between two molecules (10^{-13} to 10^{-11} sec.). The "classical picture" of a collision is that of a small change in the natural frequency of the molecule due to the influence of an approaching molecule, which causes the oscillating molecule gradually to get out of phase with the exciting radiation. This effect persists over some hundreds of oscillations throughout the duration of a collision, and the integrated effect causes such a large phase shift that the absorption or emission of radiation is interrupted. The perturbation of the energy levels caused by the approaching molecule is, however, small compared with the separation of the two levels between which transitions are being induced. This is very different from the situation in the ammonia inversion spectrum. Here a collision is caused by a large perturbation of the energy levels acting for less than one period of the radiation.

§ 3. THE CROSS-SECTION FOR COLLISIONS WITH NON-POLAR MOLECULES

Since the abnormally large collision diameters in pure ammonia are ascribed to the effects of dipole-dipole interaction, it would be expected that the cross-section for collisions between an absorbing ammonia molecule and a non-polar molecule would be considerably smaller. Measurements were therefore undertaken of the line breadth constant of the strongest line in the inversion spectrum of ammonia ($J, K=3,3$) for admixtures of six non-polar gases. Since none of the gases used had appreciable absorption at centimetre wavelengths, the effect of diluting the ammonia with one of them is to increase the collision frequency for the same partial pressure of ammonia and thus to broaden the line. Although it was possible to determine the line breadth constant for the mixture by the method already described (Bleaney and Penrose 1947) for the inversion lines of pure ammonia, the close agreement with the theoretical intensity obtained in the previous measurements on this line made it possible to use a different technique which was not only simpler but rather more accurate. Instead of measuring the absorption coefficient at a number of frequencies in order to delineate the line, the absorption coefficient was measured only at the centre of the line. For a mixture with a non-absorbing gas the absorption there is smaller than in pure ammonia, varying inversely as the line breadth constant at pressures where the absorption due to tails of neighbouring lines is negligible. The formula for the absorption coefficient at the centre of a line is

$$\alpha = \frac{4\pi^2 |\mu_{JK}|^2 \bar{\nu}_0^2 N_{JK}}{3kT\Delta\bar{\nu}} \quad \text{per cm.} \quad \dots\dots(4)$$

where $|\mu_{JK}|^2$, N_{JK} and $\Delta\bar{\nu}$ are as already defined and $\bar{\nu}_0$ is the wave number of the line in cm^{-1} .

The line breadth constant $\Delta\bar{\nu} = f/2\pi c$, where f is the collision frequency. For a given mixture both f and N_{JK} vary directly as the pressure, and the absorption coefficient at the centre of a line is therefore independent of the pressure. The addition of a non-absorbing gas to a certain concentration of ammonia increases f from the value for pure ammonia, f_{11} , to $f_{11} + f_{12}$, where f_{12} is the frequency of collisions between an ammonia molecule in the state (3,3) and molecules of the foreign gas. The absorption is therefore reduced in the ratio

$$\alpha_2/\alpha_1 = f_{11}/(f_{11} + f_{12}),$$

where α_1 is the absorption coefficient for pure ammonia and α_2 that for the mixture.

Hence $f_{12}/f_{11} = (\alpha_1/\alpha_2) - 1$.

If S_{12} and S_{11} are respectively the cross-sections for ammonia-foreign gas collisions and for ammonia-ammonia collisions, the expressions for f_{12} and for f_{11} are

$$f_{12} = 2N_2 S_{12} (2RT/\pi)^{1/2} (M_1^{-1} + M_2^{-1})^{1/2},$$

$$f_{11} = 4N_1 S_{11} (RT/\pi M_1)^{1/2},$$

where N_1 , N_2 are the numbers of ammonia and foreign gas molecules per cm^3 respectively, and M_1 , M_2 the molecular weights of ammonia and of the foreign gas.

Hence the ratio of the cross-sections is

$$\frac{S_{12}}{S_{11}} = \frac{N_1}{N_2} \frac{f_{12}}{f_{11}} \left(\frac{1}{2} (1 + M_1/M_2) \right)^{-1/2} = \frac{N_1}{N_2} \left(\frac{\alpha_1}{\alpha_2} - 1 \right) \left(\frac{1}{2} (1 + M_1/M_2) \right)^{-1/2}. \quad \dots\dots(5)$$

§ 4. EXPERIMENTAL

The six non-polar gases used in the mixtures were helium, hydrogen, nitrogen, oxygen, argon and carbon disulphide. The ratios in which the mixtures were made up were chosen so that the line breadth constant for the mixture would be approximately double that for the same partial pressure of pure ammonia. Two slightly different techniques were used to prepare the mixtures, the choice depending on whether the non-polar gas would condense in liquid air or not.

The mixtures were prepared in a bulb of about one litre volume which had a small side tube which could be immersed in liquid air. In cases where the foreign gas would not condense at liquid air temperature, the bulb was first filled with ammonia at a known pressure. After the bulb has been isolated, the ammonia was frozen out in the side tube (the vapour pressure of ammonia at 90°K . is about 10^{-6} mm. Hg). Then a suitable pressure of the foreign gas was admitted to the bulb, after which the bulb was again isolated. Finally the liquid air bath was removed, and the ammonia in the side tube allowed to evaporate and mix with the other gas, several days usually being allowed to elapse to ensure proper mixing. Since the volume of the side tube was negligible in comparison with that of the bulb, the error due to its different temperature was negligible.

When the foreign gas would condense at liquid air temperature, the following procedure was adopted. A second bulb was filled to a known pressure with the foreign gas, which was then transferred to the first bulb by freezing it into the side tube by means of a liquid air bath. This operation was repeated with ammonia. The first bulb was then isolated, and the gases allowed to evaporate and mix as before.

In each of these methods a given volume was filled to a known pressure with each of the constituents of the mixture. The ratio (N_2/N_1) of the numbers of molecules per cm^3 of the two gases was therefore accurately known.

The intensity of absorption of the mixture of gases was determined by observing the reduction in power transmitted through a cavity resonator of known Q , when the gas was admitted. A full account of this technique has already been given in paper A. In terms of the detector readings δ_0 and δ_1 before and after admitting the mixture, the absorption coefficient is given by the expression $\alpha = (2\pi/\lambda Q_0)((\delta_0/\delta_1)^{\frac{1}{2}} - 1)$. The absorption was measured for a number of pressures such that the partial pressure of ammonia lay in the range 0.2 mm. to 3 mm. Figure 2 shows the absorption-pressure curve for an ammonia-oxygen mixture in the proportion $N_1/N_2 = 0.107$. The falling-off in the absorption at low pressures is due to the disturbance of thermal equilibrium (Bleaney and Penrose 1948), while the rise at high pressures is attributable to

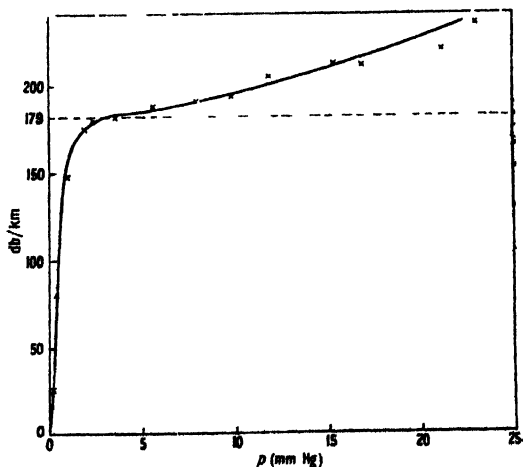


Figure 2. Absorption in ammonia-oxygen mixture at 0.7964 cm^{-1} .

the contributions from other lines in the spectrum. The shape of the full curve includes an allowance for this rise, and provides a means for smoothing the points. The dotted line represents the corrected value (179 db/km.) for the line (3,3) in this mixture. The absorption coefficient for the centre of this line in pure ammonia is 360 db/km. (see A, or table 1 of this paper). From the ratio of these two absorptions, the ratio of the collision cross-sections for ammonia-oxygen and ammonia-ammonia can be calculated by means of equation (5). Since the latter is known accurately from previous measurements by the authors, the collision diameter (σ_{12}) for ammonia-oxygen can be evaluated from the formula $\sigma_{12} = \sqrt{(S_{12}/\pi)}$.

From similar measurements and calculations for the other mixtures, the value of the collision diameter in each case was determined.

§ 5. RESULTS AND DISCUSSION

The results of these measurements are shown in table 2.

Table 2

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Helium	4.0	0.21	0.0295	17.6	2.35	3.20	3.6
Hydrogen	2.0	0.78	0.064	38.2	3.50	3.58	4.5
Nitrogen	28.0	1.72	0.205	122	6.4	4.09	5.0
Oxygen	32.0	1.51	0.123	73.5	4.85	4.02	4.9
Argon	39.9	1.74	0.111	66.3	4.6	4.04	4.0
Carbon disulphide	76.1	8.6	0.296	177	7.5	—	6.9
Ammonia	17.0	—	—	597	13.8	4.4	—

(1) Gas mixed with NH_3 ; (2) molecular weight; (3) polarizability (in $\text{cm}^3 \times 10^{-24}$); (4) S_{12}/S_{11} measured; (5) S_{12} measured (in $\text{cm}^2 \times 10^{-18}$); (6), (7), (8) collision diameters in A.; (6) measured; (7) according to kinetic theory; (8) calculated.

In the first three columns are shown the non-polar gas, its molecular weight and its polarizability. The experimental value of the ratio of the collision cross-section to that for the line (3,3) in pure ammonia is given in column (4), and the actual cross-section in column (5). The last three columns contain respectively the experimental value of the collision diameter, the kinetic theory molecular diameter and the diameter calculated by an extension of the theory of § 2 of this paper. The values of the kinetic theory diameter are not, of course, measured directly for each mixture; the figure given is obtained by taking the sum of the radii of ammonia and of the admixed gas (Kennard 1938).

The calculated diameter given in the last column is obtained as follows. A molecule of a non-polar gas becomes polarized in the electric field of an ammonia molecule, and its induced moment produces a field which acts on the permanent dipole of the ammonia molecule. It is easy to show that the energy of the latter interaction is $r^{-6}a|\mu_{JK}|^2(1+3\cos^2\theta)$, where a = polarizability of non-polar molecule; r = distance between centres of ammonia molecule and non-polar molecule; θ = angle between dipole moment of ammonia molecule and radius vector.

If the assumption is made, as before, that a collision occurs when the energy of interaction reaches a certain value W , then by an averaging process similar to that given in the Appendix, one obtains for the collision cross-section

$$S_{12} = 3.90(\mu_1^2 a / W)^{\frac{1}{2}}.$$

If the value of W necessary for a collision is assumed to be the same as that found for an ammonia-ammonia collision, i.e. 3.0×10^{-16} ergs, the value of S_{12} , and hence the collision diameter, may be calculated.

Inspection of table 2 shows that the experimental values of the collision diameter are of the same order as the kinetic theory values, as, also, are the values calculated on the basis of the theory outlined in the preceding paragraph. The latter means that at the distance between the molecules required by the purely electrical interaction for a collision, other forces come into play, and that the collision mechanism suggested is superseded by more powerful interactions. It is satisfactory, however, that the theory does not give values of the collision diameter which are much bigger than those observed.

The collision cross-section for ammonia-nitrogen appears to be exceptionally large in comparison with those for ammonia-oxygen and ammonia-argon. This was unexpected, since the polarizability and the kinetic theory diameter of each of these molecules is very nearly the same. The first measurements were made with commercial nitrogen, and gave a collision cross-section of 128×10^{-16} cm². The experiment was then repeated with a mixture made from specially pure and dry nitrogen, for which the value shown in table 2 was obtained. The two values agree, within the experimental error, and the abnormally high value seems, therefore, to be correct. There appears, however, to be no obvious explanation.

In the rotational state (3,3) the dipole moment precesses about the axis of total angular momentum at a fairly small angle, and its projection on that axis is large. Its effect in producing interactions that may cause collisions is therefore almost a maximum, corresponding to the fact that the line (3,3) in pure ammonia is one of the broadest in the inversion spectrum. The collision diameter is 14.2 Å., whereas the line (5,1) gives a collision diameter of only 8.7 Å. in pure ammonia. This wide variation between the different lines seems unlikely to occur for collisions between ammonia and non-polar molecules, since the collision diameters are approximately the same as those given by the kinetic theory. For the same reason it is unlikely that there will be any significant difference between collisions which broaden the line and those which produce thermal relaxation (Bleaney and Penrose 1948), since a kinetic theory collision must surely be effective in producing thermal relaxation.

§ 6. COMPARISON WITH OTHER CENTIMETRE WAVE SPECTRA

There are as yet few data on the widths of lines in centimetre wave spectra of molecules other than ammonia. Accurate measurements have been made on the water-vapour line at 0.742 cm⁻¹ for mixtures of water vapour and air by Autler and Becker (1946), and for pure water vapour by Townes and Merritt (1946). At a temperature of 45° C. the former workers find $\Delta\nu = 0.087 \pm 0.001$ cm⁻¹ for zero concentration of water vapour in the atmosphere, rising to 0.107 ± 0.001 cm⁻¹ at a concentration of 50 gm/m³. From these figures the value of the collision diameter for water-air is 5.4 Å., which lies between the values for ammonia-oxygen and ammonia-nitrogen given in table 2. The diameter for water-water collisions is found to be 10.2 Å., while the value calculated from a direct measurement by Townes and Merritt on pure water vapour at a pressure of 0.103 mm. Hg is 10.0 Å. These values are comparable with those for the narrower lines in pure ammonia; this is satisfactory, since the water molecule

is an asymmetrical top, and the dipole moment will have a fairly small steady component in the two rotational levels between which the centimetre wave transition takes place.

The measurements of Beringer (1946) on the oxygen spectrum at 2 cm^{-1} show that the line breadth constant at atmospheric pressure must lie between 0.02 and 0.05 cm^{-1} . This gives for the collision diameter 2.8 to 4.5 \AA ., whereas the kinetic theory diameter is 3.6 \AA . The collision diameter for oxygen-nitrogen was the same.

The main conclusion to be drawn from these results is that the collision diameter is only significantly greater than the kinetic theory diameter for encounters between permanent dipoles. For collisions between permanent dipoles and non-polar molecules the diameter is approximately the same as the kinetic theory diameter, and the measurements on the oxygen spectrum suggest that this is also true (as would be expected) for collisions between two non-electrically polar molecules.

APPENDIX

Calculation of the collision cross-section

Let $\mu_1 = \mu(K^2/(J^2 + J))^{1/2}$ be the mean dipole moment averaged over the rotation of the molecule absorbing radiation, and let μ_2 similarly be the mean dipole moment of an approaching molecule which causes an interruption of the absorption. Then the latter produces at a point whose coordinates are (r, θ) with respect to the direction of its mean dipole moment an electric field E whose magnitude is

$$|E| = \mu_2 r^{-3} (1 + 3 \cos^2 \theta)^{1/2}.$$

If the dipole μ_1 makes an angle ϕ with this field, the interaction energy is

$$\mu_1 E \cos \phi = \mu_1 \mu_2 r^{-3} \cos \phi (1 + 3 \cos^2 \theta)^{1/2}.$$

The assumption is made that a "collision" occurs when this energy reaches a certain value W . Then the dipole μ_2 can be imagined as surrounded by a surface at which the interaction with the dipole μ_1 reaches the required value for a collision. Owing to the directional nature of the dipole this surface is not spherical, but has the form of a prolate spheroid whose major axis is twice as long as the two minor axes. To find the mean cross-section of the molecule it is sufficiently accurate to find the volume of this spheroid, and to use the cross-section of a sphere of equal volume. Since the interaction energy varies as $1/r^3$, this leads to a very simple mathematical procedure.

The equation for the surface of the spheroid is

$$r^3 = \mu_1 \mu_2 \cos \phi (1 + 3 \cos^2 \theta)^{1/2} / W.$$

Hence its volume V is

$$\frac{4\pi}{3} \int_0^{\pi/2} r^3 \sin \theta d\theta = \frac{4\pi}{3} \frac{\mu_1 \mu_2 \cos \phi}{W} \int_0^{\pi/2} (1 + 3 \cos^2 \theta)^{1/2} \sin \theta d\theta = \frac{4\pi}{3} \cdot 1.38 \cdot \frac{\mu_1 \mu_2 \cos \phi}{W}$$

since the value of the integral is $1 + (\ln(2 + \sqrt{3}))/2\sqrt{3} = 1.38$.

This expression for the effective volume of an approaching molecule requires to be averaged over the mean dipole moments of all approaching molecules, and over the angle ϕ which the absorbing dipole μ_1 makes with the field of the

approaching dipole. The former averaging requires simply that μ_3 be replaced by $\bar{\mu}_2$, the average of the mean dipole moments of all the molecules present; the second averaging gives simply a factor of one-half. Hence

$$\bar{V} = \frac{2\pi}{3} \cdot 1.38 \mu_1 \bar{\mu}_2 / W.$$

Equating this volume to that of the equivalent sphere, one finds for the mean cross-section area of the approaching molecules

$$S_{12} = \pi(3\bar{V}/4\pi)^{\frac{1}{2}} = \pi(0.69\mu_1\bar{\mu}_2/W)^{\frac{1}{2}} = 2.45(\mu_1^2\bar{\mu}_2^2/W^2)^{\frac{1}{2}}.$$

ACKNOWLEDGMENTS

The authors are indebted to the Director of Physical Research, Admiralty, on whose behalf part of this work was carried out, for permission to publish. They also desire to express their acknowledgments to the Royal Commissioners for the Exhibition of 1851, whose award to one of them (R. P. Penrose) made the continuation of the work possible.

REFERENCES

- AUTLER, S. H., and BECKER, G. E., 1946, *Phys. Rev.*, **70**, 300.
 BERINGER, R., 1946, *Phys. Rev.*, **70**, 53.
 BLEANEY, B., and PENROSE, R. P., 1947, *Proc. Roy. Soc. A*, **189**, 358 ; 1948, *Proc. Phys. Soc.*, **60**, 83.
 KENNARD, E. H., 1938, *The Kinetic Theory of Gases* (New York and London : McGraw-Hill Book Co.), p. 149.
 TOWNES, C. H., and MERRITT, F. R., 1946, *Phys. Rev.*, **70**, 558.

Efficiency of Counting Systems

BY M. BLACKMAN AND J. L. MICHIELS

Imperial College, London

MS. received 10 September 1947; read 6 February 1948

ABSTRACT. A review is made of the derivation of formulae for the efficiency of counting systems, comprising in the general case a counter and amplifier, a scaling circuit and a recorder. In the case of recorders of constant resolving time no complete formula has been found in the literature, and this has been derived. A number of existing formulae are shown to be valid only for a particular type of variable recorder. A discussion is given of the use of the formulae in practical counting experiments.

§ 1

IN various branches of nuclear and atomic physics it is necessary to measure the intensity of some kind of incoming radiation, the measurement taking the form of a count of the number of particles or photons arriving within a given time. For this some form of counting system is employed. In nuclear physics a counting system normally consists of a counter and amplifier feeding a scaling circuit (hereafter termed a "scaler") which in turn actuates a recorder. The counter may be an ionization chamber, Geiger-Müller tube etc., depending upon the type of

radiation to be detected. In all cases its function is to produce an electrical impulse upon the passage through it of a particle or photon of the incoming radiation. These impulses are then fed into the remainder of the system. In such a counting system the various components have finite resolving times. That is, after they have been actuated, a certain recovery time must elapse before they can again be actuated. Because of this, a certain fraction of the incoming radiation which arrives during the recovery time will not be detected. A large number of theoretical and semi-theoretical discussions occur in the literature in which the efficiency of counting systems is treated. In the simplest of all cases, in which the counting system consists of a counter, amplifier and a simple recorder, it is usual to assume an infinitely small resolving time for all parts of the system except the recorder which has a constant resolving time τ . The efficiency E of this system is then defined as the ratio of the counts recorded in a given time to the number of particles arriving in that time. It is rather astonishing to find, even in this simple case, that there are at least three different expressions for this efficiency E to be found in the literature, namely:—

$$(a) \quad E = 1/(1 + \mu\tau) \text{ (Ruark and Brammer 1937),} \quad \dots\dots(1)$$

$$(b) \quad E = \exp(-\mu\tau) \text{ (Volz 1935, Schiff 1936),} \quad \dots\dots(2)$$

$$(c) \quad E = \{1 - \exp(-\mu\tau)\}/\mu\tau \text{ (Locher 1933).} \quad \dots\dots(3)$$

In these formulae, μ is the average number of particles arriving per second, the distribution of particles being supposed to be a random one. There are also generalizations of (a) and (b) due to Ruark and Brammer (1937) and Alaoglu and Smith (1938) for cases where the finite resolving time of the counter is taken into account and where a scaler is introduced. In view of the different versions which are current, it seems worth while to examine the problem afresh in order to find which formula is correct, and how it should be generalized.*

§ 2

In discussing the efficiency of counting systems, it is essential to adopt a clear terminology for the impulses in various parts of the systems. In this paper, the incoming radiation will be said to consist of a stream of 'particles'. The electrical impulses produced by the counter will be referred to as 'pulses' and an event which actuates the recorder will be termed a 'count'. The efficiency E of the counting system is defined as the ratio of the number of particles counted in a given interval of time to the number of particles arriving at the counter within this time. In this connection it should be remarked that it is possible for some of the incoming particles not to actuate the counter for reasons other than that it is recovering from the effect of a previous particle. In a Geiger-Müller tube, irradiated by x rays, for example, only a certain fraction of the incoming photons will eject an electron from the wall of the counter and so be capable of producing a discharge within it. This cause of inefficiency is not dealt with in these calculations which are concerned solely with the effect of finite resolving times within the system. The incoming particles are assumed to be distributed at random and arriving at an average rate μ .

§ 3

Consider first a system in which no scaling circuit is employed. The recorder is assumed to have a constant resolving time τ and the counter and amplifier to have

* Results essentially in agreement with those of the present paper have recently been derived by Jost (1947) by a very elegant method. Jost's paper, however, emphasizes the mathematical rather than the physical aspect of the problem.

a resolving time negligibly small compared with this. Suppose that N counts occur in a certain experiment of duration T . After each count the recorder is inactive for a period τ . Let p be the average number of particles which occur within a period τ after a given particle. Then, on average, this number of particles is lost at each count. Hence the total number of particles arriving during the experiment is $N + pN$ and the efficiency of the system is

$$E = N/(N + pN) = 1/(1 + p). \quad \dots\dots(4)$$

It is now necessary to calculate p . It is well known that for particles distributed at random, the probability $W(m, t)$ of m particles arriving in an interval t is

$$W(m, t) = e^{-\mu t} (\mu t)^m / m! \quad \dots\dots(5)$$

Consider a time interval τ commencing immediately after the arrival of a particle. The above formula is still valid since there is no correlation between the time of arrival of particles. Hence the probability of m further particles following the first particle within the resolving time τ is $W(m, \tau)$. In such a case the number of particles lost is m . Hence p the average number lost in time τ is

$$\begin{aligned} p &= \sum_{m=1}^{\infty} m W(m, \tau) \\ &= e^{-\mu\tau} \cdot \mu\tau \sum_{m=1}^{\infty} (\mu\tau)^{m-1} / (m-1)! = \mu\tau. \end{aligned} \quad \dots\dots(6)$$

The efficiency is therefore given by

$$E = 1/(1 + \mu\tau). \quad \dots\dots(7)$$

This formula is one of three quoted in the introduction and has been derived in various ways in the literature, for example by Ruark and Brammer (1937). The formula (2) has been derived by Schiff (1936) who employs the following argument for a recorder with a constant resolving time. The probability of no particle arriving within a time τ after a given particle is $e^{-\mu\tau}$ (see equation (5) for $m=0$). This is then put equal to the probability of counting the next particle which in turn is put equal to the efficiency. This expression gives the probability that no loss will occur, but it is not immediately related to the magnitude of the probable loss which does occur. At very small counting rates, the number of particles which arrive in an interval τ will be either 0 or 1 to a very good approximation, and in this limit the efficiency $(1 - \mu\tau)$ is correct. In all other cases it bears no relation to the true efficiency. It will be seen later, however, that this formula is correct for another type of recorder in which the resolving time is not constant. Formula (3) has been given by Locher (1933). He finds

$$p = \mu\tau - 1 + e^{-\mu\tau}. \quad \dots\dots(8)$$

He takes a series of intervals of duration τ at random, and assumes that if there is only one particle within such an interval it is counted. If there are two, one is lost and so on. With this assumption

$$p = \sum_{m=2}^{\infty} (m-1) W(m, \tau),$$

which can readily be seen to lead to (8) if τ is identified with the resolving time of the recorder. He then considers an "experiment" of duration τ in which the average number of particles arriving is $\mu\tau$ and hence deduces the average fraction lost and

the efficiency, obtaining equation (3). This result differs from the correct result given by (7), particularly for very low counting rates in which region the loss given by (3) is half that given by the correct formula. The error in Locher's argument is that he has ignored the fact that the intervals τ , having been defined as times during which the recorder is inactive, start only after a particle has been recorded. Consequently if, as in Locher's calculation, such periods are chosen at random, the first particle, which actuates the recorder, will not always occur at the beginning of the period and hence the effective resolving time assumed in the calculation is in general less than τ . Evidently, if the counting rate is very low, the probability of the first particle occurring anywhere within a period τ will be constant and the effective resolving time assumed will be $\tau/2$. If the counting rate is very high, the probability of the first particle occurring very near the beginning of the period will increase and Locher's formula should then tend to the correct one, which in fact it does.

§ 4

Consider now a counting system in which a scaler of scale-factor n is used. The scaler and all components of the system previous to the recorder are again assumed to have resolving times negligibly small compared with that of the recorder which has the constant value τ . As before, an experiment is considered in which N counts are obtained in an interval T . Let p again be the average number of particles lost during an interval τ after a count. The number of particles recorded is nN and the total number arriving is $nN + pN$. The efficiency is therefore given by

$$E = nN/(nN + pN) = n/(n + p) = 1/(1 + p/n). \quad \dots\dots(9)$$

In this procedure the number of pulses ($< n$) left in the scaler at the end of the experiment has been ignored. This is a sufficiently good approximation when N is very large compared with n . In this connection it should be noted that the formula derived for the efficiency will in any case be valid only if the number of counts is large. This follows from the fact that the number of particles lost after a count is obtained by taking a statistical average over the possible cases which can occur in a large number of counts. If only a small number of counts are taken, the loss will show fluctuations about its mean value p which may be quite large. In experimental work, it is usual to count a number of particles sufficiently large for the statistical fluctuation in this number to be small. If, however, a scaler of large scale-factor is employed, the number of counts may not be sufficiently large for the fluctuation in p , the average number of pulses lost per count to be small. In applying the expressions for the efficiency of the system, it is therefore necessary to consider both the value of p and its fluctuations.

It is now necessary to calculate p in this case. Suppose that m particles arrive at the system within a time τ after a count. If m is less than n , these are stored in the scaler and none are lost. If m is equal to or greater than n but less than $2n$, then n are lost. In general, if $qn > m \geq (q-1)n$, then $(q-1)n$ are lost. Hence the expression for p is

$$\begin{aligned} p = & n\{W(n, \tau) + W(n+1, \tau) + \dots + W(2n-1, \tau)\} \\ & + 2n\{W(2n, \tau) + W(2n+1, \tau) + \dots + W(3n-1, \tau)\} \\ & + \dots \end{aligned}$$

A function $L(n, \tau)$ can be defined by

$$L(n, \tau) = \{W(n, \tau) + W(n+1, \tau) + \dots \text{to } \infty\} \\ = e^{-\mu\tau} \{(\mu\tau)^n/n! + (\mu\tau)^{n+1}/(n+1)! + \dots\}. \quad \dots\dots(10)$$

In terms of this function

$$p = n\{L(n, \tau) - L(2n, \tau)\} + 2n\{L(2n, \tau) - L(3n, \tau)\} + \dots \\ = n\{L(n, \tau) + L(2n, \tau) + \dots\}. \quad \dots\dots(11)$$

The function $L(n, \tau)$ can be expressed in terms of tabulated functions (Molina 1942).

It is evident that if $\mu\tau/n$ is small compared with unity, the series $L(n, \tau)$ will converge very rapidly. In the case where $\mu\tau/n$ is very small, an approximate expression can be obtained from (11),

$$p = ne^{-\mu\tau} \{(\mu\tau)^n/n!\}$$

$$\text{and} \quad E = 1 - e^{-\mu\tau} \{(\mu\tau)^n/n!\} \quad \mu\tau/n \ll 1. \quad \dots\dots(12)$$

The expression (11) does not, as far as we know, appear to have been given in the literature. Lewis (1937) has given an expression for the loss in the above type of system in the region of small losses. His formula gives the fractional loss in the form of an expression which is equivalent to the function $L(n, \tau)$ instead of $p/(p+n)$. This is a valid approximation. A somewhat similar expression has been given by Alaoglu and Smith (1938) who introduced the function $L(n, \tau)$. They use it in conjunction with a different formation of the efficiency of the system. They calculate the probability that the time t in which a group of n particles follow a given particle (that is, after a count) is less than τ . This probability they find to be the function $L(n, \tau)$. The probability $G(n, \tau) = 1 - L(n, \tau)$ of the time t being greater than τ , they call the efficiency of the system. This is not a satisfactory definition of efficiency for a recorder of constant resolving time, but the formal result applies to a certain type of recorder of variable resolving time. This type of recorder is discussed in § 6.

§ 5

The system considered can now be further generalized to include the case in which the counter and amplifier have a constant resolving time σ . The recorder has a constant resolving time τ . It is assumed that the resolving time of the first stage of the scaler, if not negligible, can be incorporated in that of the counter and amplifier and that those of the succeeding stages of the scaler can be neglected. In such a system, not all the particles arriving at the counter will produce pulses which are fed into the scaler and thence to the recorder. Owing to the effect of the recovery time of the counter, the pulses which reach the recorder are no longer randomly distributed in time. Alaoglu and Smith (1938) have introduced a probability function $L^*(m, t, \sigma)$ giving the probability that the time of arrival of a group of m pulses, after a given pulse, is less than a specified time t . This function can be calculated in the following way. It is first necessary to obtain the probability that a pulse will occur at a time between t and $t+dt$ after a given pulse, there being no intervening pulses. The probability that no particle will arrive within an interval t is, for a random distribution, equal to $e^{-\mu t}$. Now the probability that no pulse will occur in a period t following a pulse is the same as the probability that no particle will arrive in the interval between σ and t , since no pulse can occur

between 0 and σ , and for any time after σ an incoming particle will produce a pulse. That is, the probability that no pulse will occur in the interval t is $e^{-\mu(t-\sigma)}$. But the probability of a particle arriving in the subsequent interval dt is μdt . Hence the probability $Q^*(t, \sigma) dt$ of a pulse occurring in the interval between t and $t + dt$ after a given pulse, with no intervening pulse is

$$Q^*(t, \sigma) dt = \begin{cases} e^{-\mu(t-\sigma)} \mu dt, & t \geq \sigma \\ 0, & t < \sigma \end{cases} \quad \dots\dots (13)$$

The function $L^*(1, t, \sigma)$ which expresses the probability that the time interval between a pulse and its successor is less than t is therefore given by

$$L^*(1, t, \sigma) = \int_0^t Q^*(t, \sigma) dt. \quad \dots\dots (14)$$

In order to calculate $L^*(m+1, t, \sigma)$ from $L^*(m, t, \sigma)$, one can proceed as follows (Alaoglu and Smith 1938). Consider a pulse occurring at a time between x and $x + dx$ after a given pulse ($x < t$), there being no intervening pulses. The probability of obtaining m pulses in a time less than $(t-x)$ is by definition $L^*(m, t-x, \sigma)$. Hence the probability of a pulse occurring at x followed by m pulses within the succeeding interval $(t-x)$ is

$$Q^*(x, \sigma) dx L^*(m, t-x, \sigma).$$

Therefore
$$L^*(m+1, t, \sigma) = \int_0^{t-m\sigma} Q^*(x, \sigma) L^*(m, t-x, \sigma) dx.$$

The upper limit in the integral is determined by the fact that it must be possible to fit at least m pulses of duration σ in the interval $(t-x)$; alternatively this can be stated in the form

$$L^*(m, t, \sigma) = 0 \text{ if } t < m\sigma.$$

It has been shown by Alaoglu and Smith, using a proof by induction, that the function L^* is given by

$$\begin{aligned} L^*(m, t, \sigma) &= \Gamma\{\mu(t-m\sigma), m\} / (m-1)! \\ &= e^{-\mu(t-m\sigma)} \left\{ \frac{[\mu(t-m\sigma)]^m}{m!} + \frac{[\mu(t-m\sigma)]^{m+1}}{(m+1)!} + \dots \right\} \end{aligned} \quad \dots\dots (15)$$

Here, as before, $\Gamma(x, m)$ is the incomplete gamma function.

Now $L^*(m, t, \sigma)$ can also be considered as the probability of m or more than m pulses occurring within the interval t following a pulse. There is, however, an upper limit M to the number of pulses which can possibly occur within a given interval owing to the finite recovery time σ following each pulse. This limit is given by $t = M\sigma + f$, $f < \sigma$. When this limit is exceeded, $L^*(m, t, \sigma)$ is zero.

Let $W^*(m, t, \sigma)$ be the probability that exactly m pulses occur in an interval t following a given pulse. Then $L^*(m, t, \sigma)$ can be written as

$$L^*(m, t, \sigma) = W^*(m, t, \sigma) + W^*(m+1, t, \sigma) + \dots + W^*(M, t, \sigma).$$

That is, the probability of m or more pulses occurring within an interval t is the sum of the probabilities of exactly $m, m+1$ etc. pulses arriving within the interval t . Hence

$$W^*(m, t, \sigma) = L^*(m, t, \sigma) - L^*(m+1, t, \sigma). \quad \dots\dots (16)$$

If m is greater than M , $L^*(m, t, \sigma)$, and thus $W^*(m, t, \sigma)$ is zero.

Let p^* be the average number of pulses lost by the recorder in an interval τ following a count. Then

$$p^* = n\{W^*(n, \tau, \sigma) + W^*(n+1, \tau, \sigma) + \dots + W^*(2n-1, \tau, \sigma)\} \\ + 2n\{W^*(2n, \tau, \sigma) + W^*(2n+1, \tau, \sigma) + \dots + W^*(3n-1, \tau, \sigma)\} \\ + \dots$$

This can be written

$$p^* = n\{L^*(n, \tau, \sigma) - L^*(2n, \tau, \sigma)\} + 2n\{L^*(2n, \tau, \sigma) - L^*(3n, \tau, \sigma)\} + \dots$$

that is $p^* = n\{L^*(n, \tau, \sigma) + L^*(2n, \tau, \sigma) + \dots + L^*(qn, \tau, \sigma)\}$, (17)

where $qn < M < (q+1)n$.

If, as before, N counts are recorded, the number of pulses leaving the amplifier is $nN + p^*N$. From (6) the average number of particles arriving during a pulse is $(1 + \mu\sigma)$. Hence the number of particles arriving during the experiment is $N(n + p^*)(1 + \mu\sigma)$. The efficiency of the counting system is therefore given by

$$E = nN / N(n + p^*)(1 + \mu\sigma) = 1 / (1 + p^*/n)(1 + \mu\sigma). \quad \text{..... (18)}$$

This formula is seen to reduce to those given for the simpler cases if first σ is put equal to zero (cf. (10) and (15)) and then n is put equal to unity. Also if τ is less than $n\sigma$, p^* is zero and the only loss is that due to the counter and amplifier. This should be so, since in this case the recorder has already recovered before another group of n pulses can have arrived from the amplifier. If $\mu(\tau - n\sigma)/n$ is much smaller than one, the expression for the efficiency can be reduced to the approximate form

$$E = (1 - e^{-\mu(\tau - n\sigma)})[\mu(\tau - n\sigma)]^n / n! / (1 + \mu\sigma). \quad \text{..... (19)}$$

In figure (1), the fraction of pulses counted $1/(1 + p^*/n)$ has been plotted as a function of $\mu(\tau - n\sigma)/n$ for various scale-factors. In computing values of p^* , use has been made of tables by Molina (1942). If σ is zero, these curves give the efficiency of the system directly. In the general case, the value obtained from the curves must be divided by $(1 + \mu\sigma)$ to give the efficiency.

Feather (1943, 1945) has also, by a different method, derived expressions for the statistical distribution in time of pulses from a counter, the results being generalized to include the case when the incident stream of particles has a non-random distribution. He has applied his formulae to the calculation of the

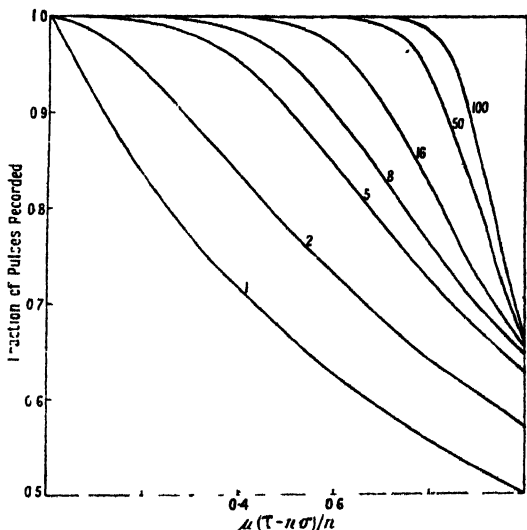


Figure 1. Fraction of pulses recorded as a function of the product of $((\tau - n\sigma)/n)$, the effective resolving time, and μ the average rate of entry of particles into the counting system, for various values of n , the scale factor of the system.

losses in a system consisting of a counter and recorder, but has not, however, extended them to take into account the effect of a scaler.

§ 6

It remains to consider the situation which arises when the resolving time of the recorder is not constant. The efficiency must then depend upon the way in which the resolving time varies, that is, upon the mechanism of the recorder. One case which arises in practice is that in which the arrival of a pulse within the resolving time re-excites the recorder without, however, recording that pulse. That is, the recovery time begins again after any such pulse. This type of recorder is termed by Ruark and Brammer (1937) a type I recorder and that with a constant resolving time a type II recorder. This is a convenient form of nomenclature. A common example of a type I recorder is the so-called Cenco counter.

Consider now a simple counting system consisting of a counter and amplifier feeding directly a type I recorder of natural resolving time τ . The resolving time of the counter and amplifier are assumed negligibly small.

Consider a stream of pulses entering the recorder. If the interval between two successive pulses is less than τ , the recorder is re-excited and the pulse lost. If the interval is greater than τ , the pulse is recorded whether the previous one has been recorded or not. Hence the probability of recording a pulse is equal to the probability, $e^{-\mu\tau}$, that the interval between two successive pulses is greater than τ . This will be equal to the fraction of pulses recorded and consequently the efficiency is

$$E = e^{-\mu\tau}. \quad \dots\dots (20)$$

This argument is due to Ruark and Brammer (1937) who attribute it to Volz (1935) and Schiff (1936).

The argument can be immediately extended to the general counting system in which the counter and amplifier have a constant resolving time σ and feed a scaler of scale-factor n actuating a type I recorder. Here the recorder registers only groups of n pulses. A group of n will be recorded only if the last pulse of the group occurs in a time greater than τ measured from the last pulse of the previous group. This is the case whether the previous group has been recorded or not. That is, if $G^*(n, \tau, \sigma)$ is the probability that the group of n pulses after a given pulse occupies an interval greater than τ , the fraction of pulses recorded is $G^*(n, \tau, \sigma)$. As before, every pulse is, on the average, associated with $(1 + \mu\sigma)$ particles.

Hence the efficiency is

$$E = nG^*(n, \tau, \sigma)/(1 + \mu\sigma) = G^*(n, \tau, \sigma)/(1 + \mu\sigma). \quad \dots\dots (21)$$

The function G^* is given by

$$G^*(n, \tau, \sigma) = 1 - L^*(n, \tau, \sigma),$$

where $L^*(n, \tau, \sigma)$ is the function considered above. That is

$$G^*(n, \tau, \sigma) = e^{-\mu(\tau - n\sigma)} \{1 + \mu(\tau - n\sigma) + \dots + [\mu(\tau - n\sigma)]^{n-1}/(n-1)!\}. \quad \dots\dots (22)$$

Equation (21) applies equally well in the case where σ is zero.

The above formula for E was given by Ruark and Brammer (1937) for $n=1$, and by Alaoglu and Smith (1938) in the general case. Ruark and Brammer explicitly derive their formula for a recorder of constant resolving time, that is, a

type II recorder, for which it certainly does not apply. Alaoglu and Smith refer to a "recorder which will not respond unless the interval between the count and its predecessor is greater than τ ". This statement applies equally well to type I and type II recorders. Their calculated efficiency, however, is valid only for a type I recorder.

§ 7

An experimental verification of the expression given by Ruark and Brammer has been published by Lifschutz and Duffendack (1938). This expression is

$$E = e^{-\mu(\tau - \sigma)} / (1 + \mu\sigma).$$

What was actually done experimentally, however, was to verify the formula in a case for which it was not derived. In view of the confusion which seems to exist, it is as well to quote the remarks of Ruark and Brammer (1937, p. 324): "If τ_r is greater than τ , and the recorder of a kind which cannot be re-excited while it is in action (type II), the efficiency of the apparatus is $f'/f = (\exp[-f(\tau_r - \tau)]) / (1 + f\tau_r)$ ". Here f'/f is our E , τ_r , τ , τ_i , σ and $f = \mu$.

In the experimental paper, the formula was interpreted as being applicable to recorders which are re-excited during recovery (type I). This is shown by the following quotation (p. 717): "In order to test the theory, the circuits used must satisfy the assumptions made in deriving the formulae. In referring to equation 1(a) it is clear that with $\tau > \sigma$ and $\sigma > 0$ equation 1(a) becomes $n_1 = n_0 \exp(-n_0\tau)$. In other words the recorder circuit used to test equation 1(a) must be a type I recorder". Equation 1(a) referred to is Ruark and Brammer's formula given above. A slightly different notation is used: n_1/n_0 , E and $n_0 \mu$.

The experimental work was carried out employing a type I recorder (a Cenco counter) and very satisfactory agreement was obtained with the theoretical formula. As shown above, the formula in fact should apply only to a type I recorder.

§ 8

In practice, experimental conditions are usually adjusted so that the counting rate is sufficiently low for losses to be small. In these circumstances the efficiencies of both type I and type II recorders are the same. In figure 2 the region of low counting losses is plotted in some detail. The ordinates in this diagram are values of $100p^*/(p^* + n)$, that is, the percentage loss of pulses in the recorder. The abscissae are rates of entry of particles into the system. Logarithmic scales are used on both axes. Curves are drawn for various combinations of selected values of the scale-factor n and the effective resolving time $(\tau - n\sigma)$.

It can be seen that for a given value of n , the curves for different values of $(\tau - n\sigma)$ are similar in shape, and can be obtained from one another by a simple displacement in abscissae. This is a consequence of the employment of logarithmic scales and the fact that the losses are functions of the product $\mu(\tau - n\sigma)$.

The losses due to the counter and amplifier can be obtained from the formula (18) for $x=0$. The total loss is the sum of those due to counter and recorder separately in the region of small losses. Inspection of figure 2 shows that the curves are all very nearly straight lines whose slopes increase with the scale-factor. The logarithmic plot emphasizes this effect rather strongly. The steepness of the curves for moderately large scale-factors does, however, indicate that where there is uncertainty in the value of $(\tau - n\sigma)$, any attempt to make a correction

for appreciable counting losses would fail. Further, any fluctuation in the rate of entry of particles will tend to introduce large errors into any such correction. For large scale-factors therefore, it would appear to be desirable to arrange working conditions so as to make recorder losses negligible (e.g. less than 0.1%). That is, the maximum counting rate employed should be less than that corresponding to the point on figure 2 at which the relevant curve cuts the axis. These precautions are not essential for scale-factors of 1 or 2, and here it may be possible to make small corrections if the parameters of the system are sufficiently accurately known.

In addition to fixing a safe counting rate, it is also necessary to fix a minimum total counting time. This must be sufficiently large to ensure both a small fluctuation in the total number of particles entering the system and a sufficient number of counts.

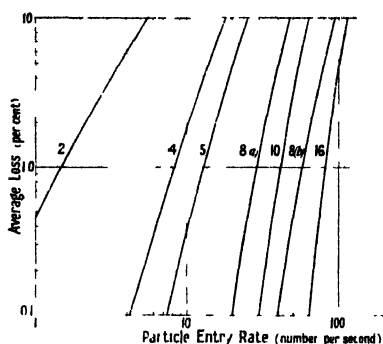


Figure 2. Logarithmic plot of the percentage loss of pulses as a function of the average rate of entry of particles into the counting system. The curves are drawn for various scale factors indicated in the diagram, with a value of 0.1 sec. for $(\tau - n\sigma)$, the effective recording time of the recorder. The two curves 8 (a), 8 (b), with values of $(\tau - n\sigma)$ equal to 0.1 and 0.05 sec. respectively, and scale factor 8, show that a change in $(\tau - n\sigma)$ results in a parallel displacement of the curve.

APPENDIX

The two tables given below summarize, in a form suitable for practical application, the data concerning the performance of various kinds of counting-devices. The authors are greatly indebted to Dr. C. E. Wynn-Williams for suggesting the form in which these tables should be compiled.

Table 1

The table gives critical mean rates of entry of particles per second for scalers of various scale factors associated with recorders of various resolving times. For entry rates less than these, the losses due to finite resolving time of the recorder are less than 0.1% and can usually be ignored in comparison with the losses due to the counter or other device at the input of the counting system and the statistical fluctuations. As the particle entry-rate is increased above these values, the fraction lost increases so rapidly, particularly for high scale factors, that it is unwise to attempt to apply a loss-correction. The working rate should therefore be kept well below the critical rate. The critical entry-rate for other values of the scale factor can be determined sufficiently accurately by interpolation on a graph of the values given in the table. Note that the critical rates are inversely proportional to the resolving times.

Recorder resolving time (seconds)	Scale factor							
	2	4	8	10	16	32	64	100
1/5	0.25	2	10	15	32	86	210	360
1/10	0.50	4	20	30	64	172	420	710
1/25	1.2	10	50	75	160	430	1000	1800
1/50	2.3	21	100	150	320	860	2100	3600
1/75	3.5	32	150	225	480	1300	3200	5300
1/100	4.5	42	200	300	640	1700	4200	7100

Table 2

The table gives maximum permissible mean rates of entry of particles per second into a scale-of-one device for different values of the mean fraction of particles lost and for various resolving times of the device. In brackets are also given the minimum time of counting (in seconds) at these maximum rates for the statistical fluctuation in the count not to exceed the instrumental losses. The table shows that it is difficult to work to an accuracy much better than about 0.5% unless long counting times are employed.

Type of device	Resolving time (sec.)	Percentage of count lost by scale-of-one device						
		0.1	0.2	0.5	1	2	5	10
Input stage of hard valve scalars limit about 10^{-7} sec.	10^{-6}	1000 (1000)	2000 (125)	5000 (8)	10,000 (1)	At higher entry-rates an adequate count is secured in less than 1 sec. but special timing devices are required.		
	2×10^{-6}		1000 (250)	2500 (16)	5000 (2)			
	5×10^{-6}		400 (625)	1000 (40)	2000 (5)			
Input stage of thyratron scalars limit about 10^{-5} sec.	10^{-5}			500 (80)	1000 (10)	2000 (2)		
	2×10^{-5}			250 (160)	500 (20)	1000 (3)		
	5×10^{-5}			100 (400)	200 (50)	400 (7)		
Geiger-Müller counters limit about 5×10^{-5} sec.	10^{-4}			50 (800)	100 (100)	200 (13)		
	2×10^{-4}	At lower entry-rates counting time must exceed 1000 sec. (about 16 min.) otherwise statistical fluctuations may exceed instrumental losses.			50 (200)	100 (26)	260 (2)	
	5×10^{-4}				20 (500)	40 (64)	100 (4)	
	10^{-3}				10 (1000)	20 (130)	50 (8)	
	2×10^{-3}					10 (260)	25 (17)	55 (2)
Modified ratchet uniselectors limit about 1.3×10^{-2} sec.	5×10^{-3}					4 (640)	10 (40)	22 (5)
	10^{-2}						5 (80)	11 (10)
	2×10^{-2}						2.5 (160)	5 (20)
Modified telephone call counters limit about 4×10^{-2} sec.	5×10^{-2}						1 (400)	2 (50)
	10^{-1}						0.5 (800)	1 (100)
		10^6	2.5×10^5	4×10^4	10^4	2.5×10^3	4×10^2	10^2
		Minimum desirable count.						

REFERENCES

- ALAOGLU, L., and SMITH, N. M., Jr., 1938, *Phys. Rev.*, **53**, 832.
 FEATHER, N., 1943, *Proc. Camb. Phil. Soc.*, **39**, 84; 1945, *Atomic Energy declassified report*, Br. 555.
 JOST, R., 1947, *Helv. Phys. Acta*, **20**, 173.
 LEWIS, W. B., 1937, *Proc. Camb. Phil. Soc.*, **33**, 549.
 LIFSCHUTZ, H., and DUFFENDACK, O. S., 1938, *Phys. Rev.*, **54**, 714.
 LOCHER, G. L., 1933, *J. Franklin Inst.*, **216**, 553.
 MOLINA, E. C., 1942, *Tables of Poisson's Exponential Limits* (New York: Van Nostrand).
 RUARK, A., and BRAMMER, F. E., 1937, *Phys. Rev.*, **52**, 322.
 SCHIFF, L. I., 1936, *Phys. Rev.*, **50**, 88.
 VOLZ, H., 1935, *Z. Phys.*, **93**, 539.

DISCUSSION

Dr. DENIS TAYLOR. The computation of counting losses is important both from the point of view of arranging the conditions of the experiment so that the losses may be reasonably small and also so that the true counting rate may be calculated from the observed counting rate. The analysis given by the authors allows this computation to be made for most of the practical cases which arise in nuclear physics. One case which is of some importance, and which seems to have escaped mention, is the case when the resolving time of the scaler itself must be allowed for. In the simple case when only the resolving time of the input stage of the scaler is important, it is presumed that the analysis given by the authors applies, but that σ , the resolving time of the detector (ion-chamber or Geiger-Müller counter), must now include the resolving time of the input stage of the scaler. Other cases can arise which are more complicated. A typical example is a pulse analyser which has been designed recently at T.R.E. This consists of a pre-sorting stage, a "scale of two", a "ring of five" and an electro-mechanical register. The resolving time of the input circuit was 100 μ sec., that of the second stage (i.e. after the "scale of two") was 2 msec., and that of the electro-mechanical register was 0.1 sec. The second stage had a high resolving time because cold cathode valves were used for the "ring of five" to economize in power supplies (30 such chains are used in the complete equipment). In a case like this, the finite resolving time of the second stage of the scaler must be allowed for. Have the authors considered extending their analysis to cover this case? We have made some vain efforts to do this at T.R.E., but eventually obtained the information we required by connecting a Geiger-Müller counter to two counting chains, one of the type described above and a second one in which the resolving time of the second stage of the scaler was sufficiently small to have a negligible effect on the counting rate. The counting losses for the second chain could be computed and, therefore, a comparison of the observed counting rates for the two chains allowed the counting losses for the former to be determined.

Another case of considerable importance is the computation of counting losses in the counting of particles from pulsed sources. The use of particle accelerators which give short pulses of particles, usually of a few seconds' duration, repeated from a few to a few hundred times a second, as, for example, with a synchrotron, frequency modulated cyclotron, etc., is likely to increase very much during the next few years, and the importance of being able to compute the counting losses in such cases cannot be overestimated. Have the authors given any attention to this problem?

Finally, I should like to call the authors' attention to a further paper on the computation of counting losses which may have escaped their attention. This is Hole, *Arkiv Mat. Astr. Fys.*, 1946, 33 A, No. 11.

Dr. G. BURNISTON BROWN. I should like to ask whether any experimental confirmation of the formula arrived at by the authors has been attempted? Physicists must not be prevented, by the fact that amplifiers etc. were mentioned, from noticing that this is a mathematical paper whose conclusion is nothing more than the consequence of assumptions made at the start and during its progress. The result would have been just as correct for elephants as for amplifiers, provided we assume that if n stimuli are given to an elephant, a fraction p of them produce a "kick", etc. Anyone making use of the theory of probability should bear in mind that it is the most obscure of all theories, and that Leibniz, the late Lord Keynes, and Dr. Harold Jeffreys all agree in admitting that they cannot define the concept of probability.

Why should modern physicists be excused from the venerable custom of *calibrating* their instruments? Arguments from the theory of probability can never take the place of experiment.

AUTHORS' reply. In reply to Dr. Taylor, we have not so far considered the extension of our calculations to the case where the second (and further) stages of the scaler have appreciable resolving times, but we think it probable that the methods used are capable of this extension. The problem of counting losses from pulsed sources is of a different type. It is, for instance, obvious that when the resolving time of the counting system is longer than the duration of a pulse only one particle can be counted per pulse. It is therefore necessary to modify even the most elementary formulae for counting losses. We are considering this

calculation in view of its general interest and are indebted to Dr. Taylor for calling our attention to this case.

With regard to the point raised by Dr. Brown, we have drawn attention in our paper to an experimental verification of one of the formulae discussed, namely the work of Lifschutz and Duffendack. It would be possible in principle to calibrate a counting system empirically if standard sources of suitable types were available in every laboratory. In some cases, however, this calibration would involve considerable difficulties, e.g. the efficiency of a Geiger counter for γ rays varies with the energy of the γ ray, and a whole range of standard γ -ray sources would be necessary.

Certain Aspects of the Mechanism of Spark Discharge

By L. B. LOEB

University of California, Berkeley, California, U.S.A.

MS. received 12 November 1947 ; read 23 April 1947

SPARKS are a class of transient occurrence in which a given existing conduction current in a gas suddenly and irreversibly changes to a current of higher magnitude, operating more efficiently by different mechanisms under conditions which rendered the previous lower current unstable (Loeb 1939, p. 408).

Certain general features of the threshold for spark breakdown, common to all mechanisms but differing in detail for the different processes, will first be presented. The more complete equation for sparks in one regime will then be used to describe the *general* properties which may be expected.

Spark breakdown in general requires fields of sufficient intensity to ensure that two processes, a primary and a secondary, are activated. The primary process is one which causes electron multiplication in the field by electron impact to reach such values that after traversing the gap of length δ the $\exp(\alpha\delta)$, or $\exp\left(\int_0^\delta \alpha dx\right)$, electrons created by one initiating electron are adequate. This is the *primary*, and predominating, process, with α , Townsend's first coefficient, representing the number of new electrons created per centimetre of path in the field direction by an electron (Loeb 1939, Chapter VIII and p. 369). Alone this process leads to field intensified currents which are proportional to the initial ionization present. The progeny of $\exp(\alpha\delta)$ electrons produced by one electron in crossing a gap is called an *electron avalanche*. The average distance traversed in the field direction to produce an ionizing act is then $1/\alpha$. The function of the secondary process is to furnish secondary electrons to replace the initiating electrons and to maintain or build up a discharge. It can take one of two possible forms (Loeb 1939, Chapter IX, pp. 377 and 403):

- (a) A secondary process *at the cathode* leading to Townsend's type of breakdown, (i) by impact on the cathode of positive ions that liberate electrons; (ii) by photoelectric effect on the cathode of photons excited in avalanche formation; (iii) by action of metastable atoms on the cathode (limited to certain gases).
- (b) A secondary process *in the gas* that leads to sparks by the streamer mechanism, i.e. photo-ionization.

To explain the behaviour observed, the liberation of electrons at the cathode by positive ion impact on the cathode is indicated as the most likely process.

As Townsend has shown, this leads to a current

$$i = i_0 [e^{\alpha \delta} / (1 - \gamma e^{\alpha \delta})], \quad \dots (1)$$

when expressed in a simplified form (Thomson 1911, Townsend 1915).

When $\gamma e^{\alpha \delta} = 1$ this expression becomes indeterminate mathematically. However, as Holst and Oosterhuis (1923) long ago indicated, $\gamma e^{\alpha \delta} = 1$ has another interpretation. This becomes clear if one reasons as follows:—

- (a) For $\gamma e^{\alpha \delta} < 1$ the discharge follows equation (1) and the current is not self-sustaining; i.e. it depends on i_0 and ceases with i_0 . It yields merely a field intensification of i_0 .
- (b) For $\gamma e^{\alpha \delta} = 1$, on the average, each avalanche of $e^{\alpha \delta}$ electrons by the primary process is multiplied by a value of γ large enough to give one new secondary electron when the $e^{\alpha \delta}$ positive ions return to the cathode. Thus one initiating electron at the cathode is able to maintain its succession indefinitely. This is then the *threshold* for a self-sustaining discharge independent of i_0 . It marks the sparking threshold.
- (c) For $\gamma e^{\alpha \delta} > 1$ the ionization of successive avalanches is cumulative and more ions are created than start. As electrons have a mobility 100 times that of positive ions, positive space charges accumulate in the gap near the cathode. These produce a very efficient secondary mechanism that allows of the discharge continuing as glow or arc at a higher current. The space charge will grow and the spark will materialize the faster the greater the amount by which $\gamma e^{\alpha \delta}$ exceeds unity.

The mechanism of such a spark then proceeds as follows. An avalanche crosses a gap of length δ at a speed of the order of 10^7 cm./sec. in sparking fields at atmospheric pressure in a time T_e . Arrived at the anode the $e^{\alpha \delta}$ electrons are absorbed and the positive ions, of which 50% are created within the last $1/\alpha$ cm., start back to the cathode. Arrived there in a time T_+ , roughly 100 times as great as T_e , these ions yield $\gamma e^{\alpha \delta}$ electrons which start a new chain of avalanches, again taking a time T_e and T_+ . At this time $\gamma^2 e^{2\alpha \delta}$ electrons begin a third sequence. When, as a result of η trips, the positive ion space-charge of $\gamma^{\eta-1} e^{\eta \alpha \delta}$ positive ions produces a self-sustaining field for a glow, the spark is complete. Its time of formation, or *formative time lag* is

$$T_f = (T_+ + T_e)\eta. \quad \dots (2)$$

If the secondary mechanism had involved photoelectric action at the cathode, the time T_f would not have involved T_+ , for the photons reach the cathode without delay. Thus T_f would be merely ηT_e . What evidence there is, however, indicates that photoelectric emission at the cathode is largely ineffective because of the wide geometrical dispersion of photons, for it will be rare indeed that a photon is liberated near the origin of the initial avalanche. Thus, in general, the γ mechanism is responsible for sparks and the value of T_f is given by equation (2).

The question has been studied by Schade (1937) using equation (2) with Townsend's mechanism active at lower pressures and has been verified as being in accord with observation. For gaps in Ne and H₂, T_f has a value of 10^{-1} to 5×10^{-5} , and has been shown to depend on T_+ and η .

Near sparking in air at atmospheric pressure the quantity $\alpha \delta$ has the relatively small value of about 17. Furthermore, ionization is a chance phenomenon and

it is not certain that every electron will ionize each time that it has traversed $1/\alpha$ cm. There is also nothing to prevent electrons by accident from ionizing twice in a distance $1/\alpha$, even though on the average it does so only once. Thus the individual avalanches will yield an electron multiplication which fluctuates above and below $e^{\alpha\delta}$ in a purely random fashion, i.e. if $\alpha\delta$ varies by ± 1 the avalanche may vary by a factor of e or $1/e$. Again electron liberation by positive ion impact is purely a chance phenomenon with an *average* value γ under any set of conditions. Thus the quantity $(\gamma e^{\alpha\delta})_i$, as yielded by *individual* avalanches will fluctuate about the *average* value $\gamma e^{\alpha\delta}$ in considerable measure.

Accordingly even though the potentials are chosen so that the average $\gamma e^{\alpha\delta} = 1$, individual avalanches occurring will yield other values of the product. Thus if V_s is the potential corresponding to $\gamma e^{\alpha\delta} = 1$, an occasional fortunate avalanche will build up a spark at a potential V less than V_s , and at a potential V greater than V_s an occasional unlucky avalanche will not give the necessary perpetuating electrons and there will be no spark. It is clear then that in individual avalanches sparks occur at values of V below and fail at values above the value of V fixed by $\gamma e^{\alpha\delta} = 1$, through the relations of α/p to X/p .

If now there be applied to a gap a potential V with initial value less than V_s and the fraction of P_s initiating electrons which give avalanches resulting in the formation of a spark be plotted against V , the curve A of figure 1 would be observed.

This curve is the integral from $V=0$ to $V=V$ of the fluctuations of the value of $(\gamma e^{\alpha\delta})_i$ for individual electrons about the average value for which $\gamma e^{\alpha\delta} = 1$ and $V = V_s$. Thus differentiation of the curve A of figure 1 gives the chance of a variation of $(\gamma e^{\alpha\delta})_i$ in a sequence of avalanches about the mean value as expressed in terms of V . The differential curve is shown in figure 1, curve B. If the derivative gives a symmetrical distribution, V_s will be the maximum value, and in the integrated curve it will represent the point of inflection at $P_s = 0.5$. With values of $\alpha\delta \sim 20$ this is nearly true.

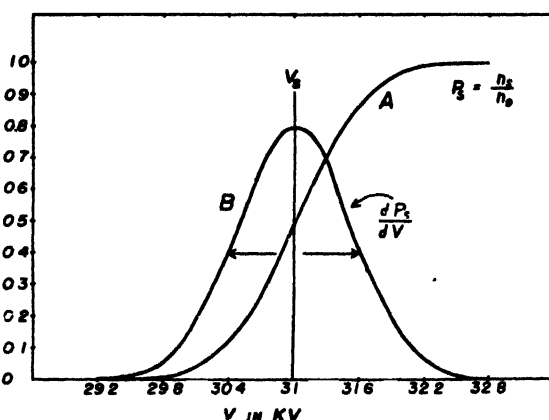


Figure 1.

- (A) Probability of sparking P_s plotted against applied potential in kilovolts.
 (B) Derivative of sparking probability against voltage indicating the distribution of sparking events about the threshold plotted in terms of voltage.

Curves of the type of curve A in figure 1 will be obtained if the fraction of individual avalanches leading to a spark are observed and plotted against V . By measuring time lags using varying but known ultraviolet illumination of the cathode and different potentials V , Wilson (1936) obtained curves analogous to figure 1 curve A and used these to evaluate V_s . The width of curve B of figure 1 depends on the magnitude of the fluctuations in γ and $e^{\alpha\delta}$ and on the number of sequences η to build up the space charge. While any one of the steps in the sequence of η successive events could interrupt the spark, it is probably only the

early ones where not many positive ions are present that are most critical. Thus at V_s where η is large the curve may be somewhat broader than at higher values of V . The width of curves calculated for Schade's data by Hertz (1937) using fluctuations of γ only lie around 1%. Actual widths may be much larger as fluctuations in ϵ^{ab} are larger than in γ . No experiments using individual electrons have been made.

Whether the distribution be wide or narrow, it is important to know that this distribution is always present and that sparks will sometimes occur below a potential V_s based on *average* values and will sometimes fail above. The value of V_s , unless the curves are very distorted, will be that at which $P_s = 0.5$. Thus, to be certain of the sparking threshold experimentally, a $(P_s - V)$ curve must be obtained from individual avalanches and V_s chosen as above. It should further be noted that T_f varies with V about V_s and that properly to interpret formative time lag studies these lags must be established relative to V_s .

When a potential V is applied, as in sparking potential measurements, it is observed that even at V_s , or above, the spark does not occur at once. The time between applications of V and the appearance of the spark is called the *time lag of sparking*. We shall call it the *observational* time lag, T_0 . In 1925 Laue and Zuber showed it to be composed of a statistical time element T_s , depending on the chance of arrival of an initiating electron and the chance that this electron caused a spark, and of a formative time element, T_f which is the time taken for the spark to grow, i.e. $T_0 = T_s + T_f$. In turn, T_s depends on $T_a = 1/n_0$, the average time between liberation of initiating electrons, which is the reciprocal of the number of electrons per second, and the chance P_s that these events will give a spark. That is, $T_s = T_a/P_s = 1/n_0 P_s$. Thus $T_0 = 1/n_0 P_s + T_f$. But by equation (2) T_f on Townsend's theory is $T_f = (T_+ + T_e)\eta$. With gap length δ , field strength X , and ion mobility K_+ , T_+ is given by $T_+ = \delta/K_+ X$, and T_e is given by $T_e = \delta/K_e X$, where K_e is the electron mobility. Thus

$$T_0 = (\delta/K_+ X + \delta/K_e X)\eta. \quad \dots\dots(3)$$

It must be also recognized that η decreases rapidly with increase of V above V_s as α/p changes rapidly with X/p . Thus for overvoltages $(V - V_s)/V_s$ which are not very large η approaches unity and T_f approaches $T_f = \delta/K_+ X + \delta/K_e X$. Hence for Townsend sparks we can write

$$T_0 = 1/n_0 P_s + (\delta/K_+ X + \delta/K_e X)\eta. \quad \dots\dots(4)$$

If now $T_a = 1/n_0$ be made small by large n_0 , and if P_s and η be made unity by using a sufficiently large overvoltage $(V - V_s)/V_s$, then T_0 becomes sensibly

$$T_0 = \delta/K_+ X, \quad \dots\dots(5)$$

since $K_0 \gg K_+$. For rare sparks below V_s observed with large n_0 , the value of T_f will be slightly larger than that of equation (3) due to the decrease in XK_+ . The conditions given by equation (5) are those under which most experimental studies of both V_s and T_0 have been made—except those whose purpose it was to study T_a . This means that true values of V_s have rarely been observed, but with large n_0 and the narrow statistical range of V about V_s the values are not seriously wrong. It also means that by using large values of $(V - V_s)/V_s$ the quantity T_0 observed has been the *characteristic* value of $T_f = \delta/K_+ X$ for the Townsend mechanism. Observations of T_0 near V_s would give exceptionally long and non-informative values of T_f . It may be noted here, however, that values of n_0 must not be too high, since various workers (see Loeb 1939, pp. 386, 448-9 Loeb and Meek 1941,

pp. 32, 172) have shown that a heavy n_0 with much field intensification can cause a space charge distortion that lowers the sparking potential or of itself leads to a spark.

It may also be noted that n_0 can be determined, even at breakdown fields, by measuring the value of the photoelectric current at two potentials below cumulative ionization and evaluating the constants in J. J. Thomson's equation for photo-ionization currents in gases (Loeb 1939, Chapter VII, Sect. 3, p. 310 and equation, p. 313; Johnson 1948). This enables the value of the current from the cathode, n_0 , to be calculated for purposes of study.

It must finally be concluded that to evaluate V_s one must observe T_0 under the following conditions: T_f must be kept much less than T_s and T_s is increased by a controlled n_0 which must be sufficiently small. Then sensibly $T_0 = 1/n_0 P_s$ and by observing T_0 for different values of V about V_s , P_s can be evaluated as a function of V and by differentiation the peak value will locate V_s .

Practically all studies except Schade's have so far been made on sparks passing by the streamer mechanism and not by the Townsend mechanism. It is thus essential to see how the considerations above apply to the streamer.

The streamer starts with an avalanche of $e^{\alpha x}$ electrons where x is nearly equal to the gap length δ at V_s . If $e^{\alpha x}$ is great enough, electrons of the avalanche passing to the anode leave behind an adequate positive space charge to initiate a streamer. Meek (1940) and Raether (1941) independently assessed this as occurring when the space charge tip field X' equalled the imposed field X . The density is determined by $e^{\alpha x}$ and the radius of the avalanche head ρ produced by diffusion of charges during advance. The diffusion is restricted in scope so that $e^{\alpha x}$ and the pressure primarily determine the density of charge. To enable a streamer to advance, the space charge tip field plus the imposed sparking field must be great enough over an adequate small volume element $\Delta x'$ deep and solid angular aperture $\pi/2$, so that a photoelectron produced in this volume in advancing to the tip can extend the space charge by making $\exp\left(\int_{\Delta x'}^0 \alpha' dx\right) = \exp\left(\int_0^x \alpha dx\right)$. Here α is the value in the undistorted field X where the electron avalanche travels a distance x to start a streamer and α' is the much higher value in the vector field $X + X'$ between $\Delta x'$ and ρ . If several photoelectrons can be produced in $\Delta x'$ the volume can be smaller or X' less. The production of photoelectrons in a distance $\Delta x'$ depends on the density of photon production in the avalanche tip and on the absorption coefficient μ of the gas for wavelengths which produce photo-ionization, as well as on $\Delta x'$. Photon production varies directly with the number of electrons $e^{\alpha x}$ and can be set as $fe^{\alpha x}$, where f is a nearly constant multiplier which may be greater or less than unity, but probably less. It should be noted in passing that the density of the molecules in these mechanisms will play a very critical rôle since ρ and μ are pressure dependent, so that pressure will be an important parameter, apart from the rôle of X/p .

It is now clear that while the avalanche formation is common to both Townsend and streamer mechanisms so that only one initiating electron is needed in both, the probability γ of the Townsend mechanism is replaced by the chance of photo-ionization in restricted volumes $\Delta x'$ in the streamer mechanism which might be lumped in a factor ϵ . Thus streamer progress likewise depends on $fe^{\alpha x}$ for photon production, multiplied by a probability factor ϵ akin to γ , which is now complex and involves μ , $\Delta x'$ and the criterion of Meek and Raether. Thus one may set the

sparkling threshold by streamer mechanism as $\epsilon f e^{\alpha x} = 1$ for streamer propagation. It is at once clear that with streamers the statistical time lag T_s will, by its dependence upon n_0 and P_s , play the same rôle as with Townsend's mechanism except that now P_s will depend on other factors lumped in ϵ such as μ , $\Delta x'$, and pressure. T_f will show differences with streamer formation. As, however, the condition $\epsilon f e^{\alpha x} = 1$ is radically different from $\gamma e^{\alpha b} = 1$, the analogy is largely formal.

It was the early discovery, 1923-27, that the value of T_f for higher pressure sparks is much less than is possible on Townsend's picture (Loeb 1939, pp. 32-33, Loeb and Meek 1941, p. 449) that, together with the help of corona studies (Loeb 1939, pp. 433, 440-450) and cloud track pictures (Raether 1935, Flegler and Raether, 1936, 1938, 1940), led to the streamer theory. In this case, the time scales indicated that T_f depended primarily on the time of electron crossing, T_e . For once the electron travels a distance x , to near the anode, the streamer forms and progresses back to the cathode with a velocity several times that of the electron. This high speed of the streamer advance results from two conditions. The enhanced tip fields of positive and negative space charges in the streamer increase electron velocities, and photo-ionization in advance of both space charges leads to an increased velocity, because photons move with the velocity of light. These increases are of the order of the ratio of the square roots of the enhanced tip fields X' to the applied field X , from $\sqrt{2}$ to $\sqrt{10}$ according to Raether's computations (Raether 1940), and of the order of the ratio $(\Delta x' + \rho)/\Delta x'$, where ρ is the radius of the avalanche tip and $\Delta x'$ the width of the sensitive part. These factors can increase the velocity of the streamer by a possible factor of 2 to 10 fold above that of the avalanche advance. They depend critically on pressure and on applied field strength. Accordingly for streamer advance $T_f = x/XK_e + x/bXK_e$, where b is the multiplication factor $\sqrt{(X'/X) \cdot (\Delta x' + \rho)/\Delta x'}$. The final stage of spark breakdown after the streamer has crossed the gap is very short, for it is determined by the velocity of re-ionization of an ionized streamer channel by the steep potential wave proceeding up the streamer channel from the instant of junction of streamer head and cathode. The velocities, as shown by Snoddy, Dietrich and Beams (1937), by Allibone and Meek (1938 a, b), and by Schonland *et al.* (1934, 1935, 1937) are of the order of 10^{10} cm/sec. The time lag of the return stroke is T_p . Thus for a streamer,

$$T_f = x/XK_e + x/bXK_e + T_p, \quad \dots\dots(6)$$

Since $b > 1$ and T_p is very small, one has sensibly $T_f \approx x/XK_e$. Here again with increased potential, i.e. overvoltages, b is greatly increased. Furthermore, with overvoltages streamers form in mid-gap at a distance x from the cathode much less than δ as α increases rapidly with V . Such streamers have been observed with Kerr cell shutters by Dunnington (1931) and White (1934) in overcharged gaps. Hence, T_f will decrease with $(V - V_s)/V_s$ very rapidly and continuously in the proportion x/δ .

Thus with the streamer mechanism statistical and formative time lags appear. There is, however, one difference apart from the fact that the value of T_f for streamers is much less than that for Townsend's discharge ($\delta/X_s K_e \ll \delta/X_s K_+$). This is that T_f has a fairly sharply defined *upper* limit, $\delta/XK_e + \delta/bXK_e + T_p$ for streamers, and *decreases* rapidly with V above V_s ; in a Townsend discharge on the other hand T_f is *indefinitely great* at V_s and, for slight increases in V above V_s , *decreases* rapidly towards a *minimum* value δ/XK_+ , the characteristic value.

In a further development of the streamer theory of spark discharge, one of the critical data yet to be determined is the condition for the change over from a Townsend to a streamer mechanism (Loeb and Meek 1941, pp. 49, 71, 76). From a knowledge of this much information on streamer propagation could be derived; it is hoped that some of this lacking information may be provided by the researches now being carried out by Meek and Craggs. It is indispensable for sparking theory, however, that the transition be observed as a function of pressure and gap length. The only way in which this can be done today is by means of a careful study of time lag, thus establishing the value of P_s and the position of V_s from a knowledge of T_0 and n_0 . Then by increasing n_0 so that $1/n_0 P_s \ll T_f$ one can observe T_0 when $T_0 = T_f$, and so determine T_f as a function of $(V - V_0)/V_s$. Then from the value of T_f observed in relation to δ/XK_+ and δ/XK_0 the operative mechanism is clear. This must be done with δ constant for different pressures p and with p constant with different gap lengths δ .

That such transitions will not occur at a single value of the product $p\delta$ for a given gas is indicated by the fact that unlike the Townsend mechanism, neither the streamer mechanism nor a pure space charge breakdown follow Paschen's law (Loeb 1939, p. 410, Loeb and Meek 1941, pp. 8, 46, 72, 81, Varney *et al.* 1935) which states that $V_s = f(p\delta)$. Physically interpreted, the failure of Paschen's law and the principle of similitude with streamers follows from the fact that a Paschen's law breakdown depends on the number of ions created in the gap, while the streamer mechanism depends on ion concentrations. The failure of Paschen's law has not been observed at ordinary pressures, as the present day accuracy of measurements of V_s with the peculiar mathematical character of the sparking equations, either Townsend or streamer, are not adequate. Thus characteristic data in the transition region are urgently called for. It is essential, however, at this point to study the character of the departure from Paschen's law in streamer formation as shown by Meek's theory alone in order to proceed further.* Meek's equation may be written as

$$f\left(\frac{V_s}{p\delta}\right)(p\delta) + \ln f\left(\frac{V_s}{p\delta}\right) = 14.46 \ln \frac{V_s}{p\delta} - \ln p\delta + \ln \delta, \quad \dots\dots(7)$$

where $V_s = X_s\delta$ for a uniform gap. Here it is seen that V_s is a function of $p\delta$ as Paschen's law predicts up to the last term $\ln \delta$. The Meek equation again takes the form,

$$\alpha\delta + \ln \alpha/p = 14.46 + \ln X/p + \frac{1}{2} \ln \delta/p \quad \dots\dots(8)$$

with
$$\frac{\alpha}{p} = f\left(\frac{X}{p}\right) = f\left(\frac{X\delta}{p\delta}\right) = f\left(\frac{V}{p\delta}\right). \quad \dots\dots(9)$$

The form of the function $f(X/p)$ for lower values of X/p in the streamer region is either

$$\alpha/p = A \exp(BX/p), \quad \text{or} \quad \alpha/p = A((X/p) + B)^2, \quad \dots\dots(10)$$

the latter holding for the higher range of X/p . What is done in solution of Meek's equation for V_s by trial and error is as follows: For a given p and δ a value of X/p is chosen and the corresponding value of α/p obtained from the curve for $\alpha/p = f(X/p)$. The values are then inserted into equation (8) and the relative

* It must be noted, however, that apart from the concentration effects in Meek's equation the value of μ is also independently pressure dependent so that the pressure dependence may be much greater than calculated above.

magnitudes of the two sides of the equation observed. A value of X/p is next chosen to make the other side of the equation larger. Thus by successive approximations, or by the use of Simpson's rule, a value of X/p can be obtained to satisfy the equation. This value at X/p is X_s/p and yields V_s from $X_s\delta = V_s$. It is at once seen from equation (8) that if one calculates V_s for δ constant and p varying or for p constant and δ varying, different values of V_s will be obtained for the same product $p\delta$. This is, of course, because if δ is increased and p decreased, $\ln(\delta/p)$ increases, while for the same $p\delta$ if p is increased and δ decreased, $\ln(\delta/p)$ decreases. The changes however, are small because of the logarithmic function. Therefore, for any one value of $p\delta$ there will be two values of X_s/p and hence of V_s depending on whether the value $p\delta$ is reached by decreasing δ and increasing p or vice versa. The difference between the two values will be the greater the greater the range of variation in δ and p , but it will be relatively small in its effect on V_s , for V_s depends on α (equations (9) and (10)) which varies rapidly with X_s , while compensating changes in δ/p in equation (8) would have to be very great since they alter the left-hand side by only $\ln \delta/p$. Experimentally, therefore, it would be only for larger ranges of variation that deviations from Paschen's law would be detected experimentally.

Failure to recognize this fact has in the past led to misunderstanding. When Meek first studied his equation he put it to test by calculating V_s as a function of $p\delta$, using the value of δ at 760 mm. Hg initially and decreasing p . In consequence, he obtained a V_s curve in which $\ln(\delta/p)$ increased as p decreased. This required a larger value on the right-hand side of the equation with the result that he plotted a $(V_s, p\delta)$ curve which was high. He noted a marked departure of the $(V_s, p\delta)$ curve from the experimentally observed values of Whitehead, at $p \sim 200$ and *assumed* that this was the point at which Townsend's mechanism changed to streamer mechanism (Meek 1940, Loeb and Meek 1940, 1941, p. 49). Roughly Dunnington's spark photographs with Kerr cell shutter also indicated a similar value as the point of appearance of mid-gap streamers in an overcharged gap (Dunnington 1931). Actually the appearance of mid-gap streamers depends not on the product $p\delta$ alone but also on p and/or δ . Later L. H. Fisher (1946) computed V_s as a function of $p\delta$ from Meek's equation using constant p and varying δ . He thus found a curve which started at the same δ as Meek's curve for atmospheric pressure $p = 760$, but which began to fall below Meek's curve as $p\delta$ decreased. This curve appeared not to depart seriously from the observed

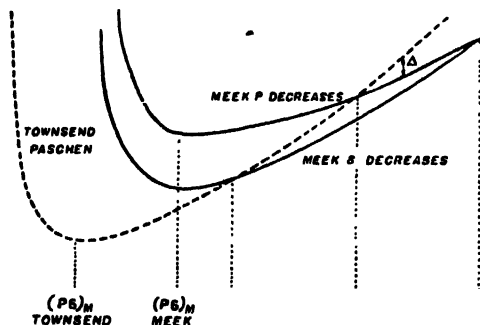


Figure 2. Represents the sparking potential plotted against product $p\delta$ for Townsend's equation and Meek's equation. Note, that starting from a value $p_0\delta_0$, Meek's equation has two values, depending on whether δ is kept constant and p decreases or p is constant and δ decreases (solid curves). The dotted curve is Townsend's equation, following Paschen's law.

data taken from Schumann's book (1923). These data are, however, variable in their reliability, and it is necessary to take into account the method by which they were obtained. Presumably, they were obtained with decreasing δ to conform to Fisher's calculation. Thus Fisher noted no departure from observed curves, even at values where streamers must have given way to Townsend's mechanism. This precipitated a rather complicated situation whose interpretation may be clarified by reference to the curve shown in exaggerated form in figure 2. The chosen point of departure for calculation is $p_0\delta_0$. D_p is the point at which the Townsend mechanism and Paschen's law take control for p decreasing. Meek's theory no longer holds. D_s is the point at which Townsend's discharge and Paschen's law take control for δ decreasing. Meek's law no longer holds. The dotted curve is the Townsend curve with a γ following Paschen's law and represents the observed curve up to streamer formation at the two different transition points.

The minima $(p\delta)_m$ for Meek's two curves fall at the same value of $p\delta$ in both cases as they reflect merely the point where $\alpha/p = f(X/p)$ passes through its point of inflection. The value of $(p\delta)_m$ will be lower for Townsend's theory as the quantity γ is increasing markedly with X/p in this region of values of $p\delta$ and will certainly be more effective than the *photo-ionization in the gas*.

With this understanding of what may be expected, one can say that the differences in the calculations of Meek and Fisher indicate the character of the departure from Paschen's law. None of the data which they use are of any value in this region of $p\delta$ to indicate this and thus accurately to confirm theory.

Fortunately, recent results of Skilling and Brenner (1941), and particularly of Howell (1939) and of Trump, Stafford and Cloud (1941) with a large Van de Graaff generator have unequivocally shown the failure of Paschen's law. Mr. C. G. Miller has kindly recalculated and plotted some of their data to show this. Curves of V_s plotted against p for constant $p\delta$ should be horizontal if Paschen's law holds. The curves of Trump, Stafford and Cloud (figure 3) show that with increasing $p\delta$ and increasing V_s there is a marked and increasing slope to the curves. This linear departure is difficult to correlate with the Meek equations (7) and (8) above. If, however, we calculate δ from the data of these curves and plot V_s against δ for constant $p\delta$, we obtain the curves shown in figure 4. Here it is seen that with $p\delta$ constant and increasing δ , V_s increases with δ . This is just what equation

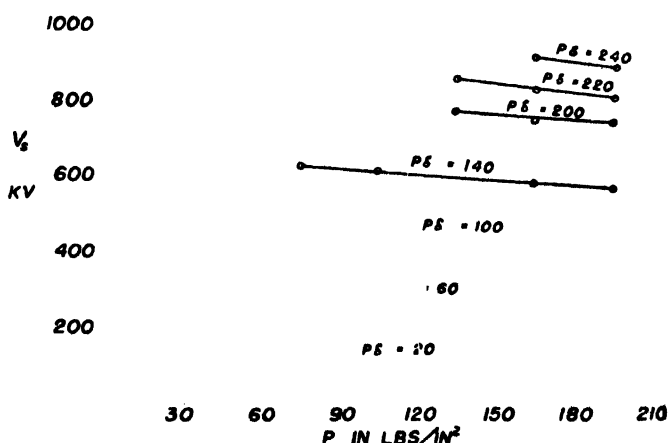


Figure 3. Curves of Trump, Stafford and Cloud showing sparking potential in kv. plotted against pressure in pounds per square inch for constant values of $p\delta$. If Paschen's law held strictly, the indicated curves for $p\delta$ would be straight lines parallel to the axis of abscissae.

(7) indicates.* Here, then, we have our first indication of the deviation from Paschen's law predicted by the streamer theory, and it is observed in regions where p and δ are varied over large ranges of values.

It should be added that in Howell's data with very high p and high V_n , above 400 kv. from the gassy surface used, the appearance of field emission of electrons from the cathode could lead to more efficient breakdown mechanisms and lowered V_n . All Howell's data at above $p\delta = 35$ have such fields and since these are too large they distorted his gap and possibly lowered V_n . The character of the curves

reported and the consistency of the data with those of Trump (below field emission) indicate, however, no departure ascribable to field emission in this region. This is doubtless due to Howell's experimental care and choice of results. The findings thus add some urgently needed confirmation to the streamer theory and the use of Meek's equation.

There is, however, a definite limitation to be expected in the application of Meek's equation beyond that occurring at low $p\delta$ and p

where Townsend's mechanism is more efficient. This applies particularly to large values of $p\delta$, where p remains relatively low, i.e. to long sparks at atmospheric pressure. The limitation will also appear when somewhat lower pressures and longer gaps are considered, i.e. for V_n curves where the initial value of δ is large and p is decreased. Meek's present equation takes no account of the effect of lowering the density of photo-ionization through either the reduction of α at low X/p or increased diffusion (Loeb and Meek 1941, pp. 49, 71, 76). Further, both density of ionization and absorption in the sensitive zone $\Delta x'$ rapidly decrease as p decreases. Thus before Townsend's mechanism at low $p\delta$ would normally be more effective than streamers according to Meek's theory, streamers could cease to form because of factors not included in the theory and sparks will occur on the Townsend mechanism. Should this factor cause transition as p is decreased or at low p as δ is increased then the $(V_n, p\delta)$ curve would show a slight increase in V_n with scattering of data and lowering of $p\delta$ as indicated by Δ in figure 2. Experimental data are as yet too crude to show this.

* Recent calculation (Miller and Loeb 1947) indicates that in fact Meek's theory predicts the sort of departure observed. It is, however, smaller than the observed departure (1% as against 14%). This is because the pressure sensitive effect present in ϵ is not included in Meek's equation to its full extent.

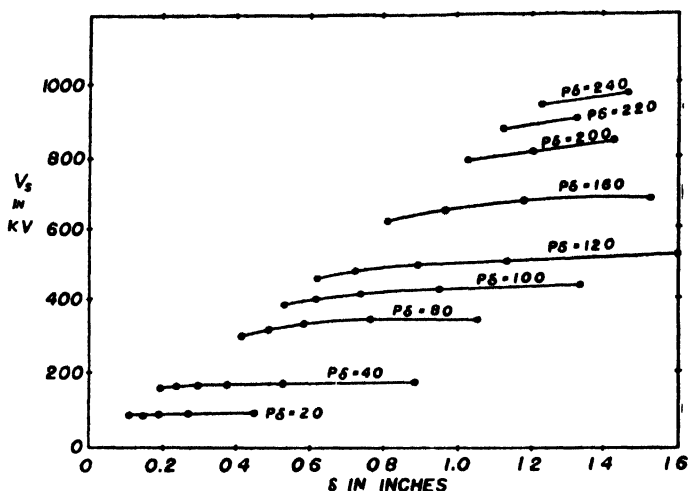


Figure 4. Curves of Trump, Stafford and Cloud as calculated by C. G. Miller with sparking potentials in kilovolts against gap length in inches for various value of $p\delta$ for comparison with Meek's equation. Meek's equation will give analogous curves, but the variation will be distinctly less than observed as indicated by later consideration of Meek's sparking condition.

It is much more important, however, to consider what happens from this circumstance in very long gaps and thus also in lightning discharge (Loeb and Meek 1941, p. 79). If either Townsend's or Meek's equation be studied it is clear that except for slowly varying logarithmic terms, the chief variation is given by $f(V_s/p\delta)p\delta$ which is nearly constant. It yields $V_s/p\delta = f^{-1}(1/p\delta)$. This indicates a hyperbolic type of decrease of $V_s/p\delta$ with increase in $p\delta$. Hence, since $V_s/\delta = X_s$, the sparking field strength, one obtains $X_s/p = f^{-1}(1/p\delta)$. Thus, the important sparking parameter X_s/p should decrease approximately hyperbolically with increasing $p\delta$. In fact, from what data we have (taken mostly at atmospheric pressure with increasing δ), up to $p\delta = 10^4$ mm. \times cm., this relation is observed as seen in figure 5. Carrying the data up to 3×10^4 mm. \times cm. from the data of Howell and of Trump, Stafford and Cloud, the value X_s/p decreases nearly linearly, within observational error, as p increases to 40 atmospheres. Correlation of the data with $p\delta$, as required by the argument above, is not very good but there is a similar trend with large scatter. In this region, however, Paschen's law deviations are extreme and the scatter with $p\delta$ is caused by the failure of this law in keeping with Meek's equation. For lower pressure regions where Paschen's law deviations are not serious, the value of X_s/p is 33 with $X_s = 25\,000$ v/cm. At 40 atmospheres, the value X_s/p is 21 with $X_s = 650\,000$ v/cm.

The cause for this behaviour is seen if one writes the Meek positive space charge field equation in the exponential form,

$$\frac{X}{p} = \frac{A(\alpha/p) \exp \{(\alpha/p)p\delta\}}{(\delta/p)^{1/2}} \quad \dots\dots(11)$$

This indicates that, neglecting the non-exponential factors, the equality, and thus the condition for sparking, is maintained if the value of α/p decreases in proportion to increase of $p\delta$. Thus, by increasing δ at p constant as in the conventional studies, α/p can decrease, and hence the value of X_s/p needed can decrease and yet give a spark by Meek's equation. This, interpreted physically, means that with increasing gap length owing to cumulative ionization the number of ions formed per centimetre of length to give the space charge field can decrease and yet yield a spark in a gap of length δ . It is true that α/p as a function of X/p increases rapidly with X/p , so that the decrease in α is not linear with δ in the exponential term. None the less, it is there. Hence, as δ or $p\delta$ increases, $1/\alpha$, the mean distance for ionization in the avalanche, increases. This will have no significance on Meek's theory as formulated. However, as stated, this theory regards *only the space charge field* which, from the slow increase of channel radius due to diffusion, should always be adequate under equation (8). It does not regard the density of photo-ionization (Loeb and Meek 1941, pp. 49, 71, 76). As,

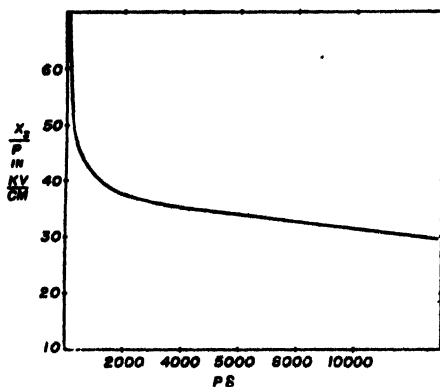


Figure 5. Sparking field strength divided by pressure in kv/cm. plotted against the product $p\delta$ in millimetres times centimetres obtained from various sparking data below the critical gap length δ_c , indicating the decline in sparking field strength with increasing gap length at constant pressure owing to increased ionization.

however, $1/\alpha$ increases, that important density decreases. Thus the photo-ionization in advance of a streamer tip can fall below the critical value if δ is increased at constant pressure. This will prohibit the spark occurring at the values of X predicted by Meek's theory. On the other hand, if path length, and pressure or gas density, i.e. $p\delta$, increase with increasing p , Meek's theory should apply.

It is thus expected to be that in long sparks and in lightning at atmospheric pressure at some critical path length δ_0 , Meek's equation should break down (Loeb and Meek 1941, p. 79). What this critical length is we do not know, but Edwards and Smee (1938), with alternating 60 c/s. potentials, observe that at gap lengths between 15 and 20 cm. at 760 mm. Hg the $(V_s, p\delta)$ curves pass through a point of inflection, increase and are different in trend for higher values of V . This is what might be expected on the anticipated breakdown of Meek's theory. For an avalanche crossing a gap greater than some critical length δ_0 can no longer produce the photo-ionization necessary, and sparks will not pass until X_s is increased to a point where $1/\alpha$ has the values required to give the needed photo-ionization. Thus, beyond δ_0 , X_s/p cannot decrease further and X_s will remain at a constant value X_{δ_0} corresponding to the limiting field at δ_0 . Sparking will then occur only at or above a field X_{δ_0} , and V_s will be given by $X_{\delta_0}\delta$ above δ_0 , irrespective of δ . The sparking potential curve given by Meek's equation will follow the full curve of figure 6. At δ_0 Meek's equation no longer holds and sparks will pass at a constant field X_{δ_0} such that $V_s = X_{\delta_0}\delta$, which increases linearly with δ as shown by the straight dashed line which is extrapolated back through the origin; the actual transition will not be sharp. This, then, indicates the trend observed by Edwards and Smee.

What actually happens in this region is that when $V_s = X_{\delta_0}\delta$ an avalanche from the cathode proceeds a distance δ_0 where its streamer starts; at the same time, with resultant field distortion, its electrons pass to the anode, initiating at a distance δ_0 or less a new streamer directed towards the cathode, and so on. At path lengths $\delta_0 \sim 20$ cm. the streamer channel will be some 0.2 to 0.5 cm. in radius, which facilitates ionization according to Raether's calculations. The statistical fluctuations in photo-ionization will smooth the range of critical potentials, fields and path lengths about δ_0 so that the transition is not sharp. The sparking potential curves should then appear as in figure 6.

Considerations of the effect of pressure indicate that δ_0 will be a critical function of pressure and rough calculations by Loeb (Loeb and Meek 1941, p. 81) placed δ_0 at 3.8 cm. at half an atmosphere. L. H. Fisher (1947), in Loeb's laboratory,

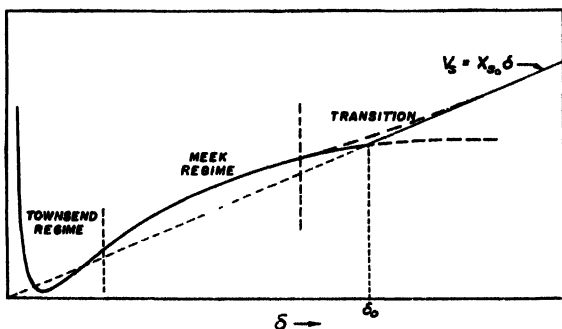


Figure 6. Sparking potential plotting against gap length at constant pressure, the full curve indicating the actual sparking potentials to be expected through the whole régime. Note the transition at δ_0 , which extrapolates back through the origin. The dot-dash curve between the two indicates the probable shape of observed sparking potential in the transition region. The Townsend régime holds for the shorter gaps, Meek's equation for the intermediate, and at longer gap lengths the sparking potential is a constant field strength times the gap length.

attempted to observe this in a carefully controlled chamber during the War. His investigation failed owing to field distortion and corona just below the critical gap length. It is urgent that this study be undertaken in some laboratory where high potentials and adequately large spark discharge chambers are available. In carrying out the study extraordinary care must be taken to avoid field distortion. The urgency of the problem will be understood when it is realized that it is the condition that determines flash-over for long static sparks and establishes lightning discharge breakdown fields.

REFERENCES

- ALLIBONE, T. E., and MEEK, J. M., 1938, *Proc. Roy. Soc. A*, **166**, 97, **169**, 246.
 DUNNINGTON, F. G., 1931, *Phys. Rev.*, **38**, 1535.
 EDWARDS, F. S., and SMEE J. F., 1938, *J. Instn. Elect. Engrs.*, **82**, 659.
 FISHER, L. H., 1946, *Phys. Rev.*, **69**, 530 ; 1947, *Ibid.*, **72**, 423.
 FLUGLER, E., and RAETHER, H., 1936, *Z. Phys.*, **99**, 625, 103, 315 ; 1938, *Ibid.*, **110**, 611 ; 1940, *Arch. Elektrotech.*, **34**, 49.
 HERTZ, G., 1937, *Z. Phys.*, **106**, 102.
 HOLST, G., and OOSTERHUIS, 1923, *Phil. Mag.*, **46**, 1117.
 HOWELL, H. H., 1939, *Trans. Amer. Inst. Elect. Engrs.*, **58**, 193.
 JOHNSON, GERALD W., 1948, "Electron Multiplication in Divergent Fields", *Phys. Rev.*, **73** (in press).
 LAUE, M., 1925, *Ann. Phys., Lpz.*, **76**, 261.
 LOEB, L. B., 1939, *Fundamental Processes of Electrical Discharge in Gases* (New York : John Wiley and Sons).
 LOEB, L. B., and MEEK, J. M., 1940, *J. Appl. Phys.*, **11**, 438, 459 ; 1941, *The Mechanism of the Electric Spark* (Stanford Press, U.S.A.).
 MEEK, J. M., 1940, *Phys. Rev.*, **57**, 722.
 MILLER, C. G., and LOEB, L. B., 1948, *Phys. Rev.*, **73**, 84.
 RAETHER, H., 1935, *Z. Phys.*, **94**, 567 ; 1940, *Arch. Elektrotech.*, **34**, 49 ; 1941, *Z. Phys.*, **117**, 386, 524.
 SCHADE, R., 1937, *Z. Phys.*, **104**, 487.
 SCHONLAND, B. F. J., *et al.*, 1934, *Proc. Roy. Soc. A*, **143**, 654 ; 1935, *Ibid.*, **152**, 595 ; 1937, *Ibid.*, **162**, 175 ; 1938, *Ibid.*, **164**, 132.
 SCHUMANN, W. O., 1923, *Elektrische Durchbruchfeldstärke von Gasen* (Springer).
 SKILLING, H. H., and BRENNER, W. C., 1941, *Trans. Amer. Inst. Elect. Engrs.*, **60**, 112.
 SNODDY, L. B., DIETRICH, J. R., and BEAMS, J. W., 1937, *Phys. Rev.*, **52**, 739.
 THOMSON, J. J., 1911, *Conduction of Electricity Through Gases*, 2nd Ed. (Cambridge : University Press), p. 490.
 TOWNSEND, J. S., 1915, *Electricity in Gases* (Oxford : University Press), p. 331.
 TRUMP, J. G., STAFFORD, F. J., and CLOUD, R. W., 1941, *Trans. Amer. Inst. Elect. Engrs.*, **60**, 132.
 VARNEY, R. N., LOEB, L. B., WHITE, H. J., and POSIN, D. Q., 1935, *Phys. Rev.*, **48**, 818.
 WHITE, H. J., 1934, *Phys. Rev.*, **46**, 99.
 WILSON, R. R., 1936, *Phys. Rev.*, **49**, 1082.
 ZUBER, K., 1925, *Ann. Phys., Lpz.*, **76**, 231.

The Electron Trap Mechanism of Luminescence in Sulphide and Silicate Phosphors

By G. F. J. GARLICK and A. F. GIBSON

Department of Physics, University of Birmingham

MS. received 30 August 1947; read 12 March 1948

ABSTRACT. Phosphorescence and thermoluminescence emission from photoconducting impurity activated phosphors have been satisfactorily explained by the storage of electrons, freed from luminescence centres or other atoms of the solid, in metastable energy levels known as electron traps. Electrons escaping from these traps give rise to emission when they recombine with luminescence centres but there is a probability that they may be re-trapped in empty electron traps before their final recombination with centres. The present theoretical and experimental studies attempt to determine the extent to which retrapping does occur and what effects it will have in modifying the phosphorescence and thermoluminescence characteristics. Theoretical treatment shows that there are marked differences in these characteristics for conditions when the retrapping process is present and for those when it is negligible. Experimental investigations of the characteristics of specimens of zinc sulphide, zinc silicate and strontium silicate phosphors indicate that, except under special conditions, retrapping of electrons is negligible. These results together with other work can be explained theoretically if it is assumed that electron traps operative in the luminescence process are spatially associated with the immediate neighbourhood of the luminescence centres formed by activating impurities. This new concept is also supported by the relations found between the luminescence characteristics and the dielectric changes in phosphors of the zinc sulphide type.

§ 1. INTRODUCTION

IT is now generally accepted that the phosphorescence and thermoluminescence characteristics of such phosphors as zinc sulphide and zinc silicate activated by suitable specific impurities are governed by the storage mechanism of electron traps which are thermally metastable. Previous workers have developed theoretical models of a luminescent solid which interpret many of the luminescence properties in terms of the trapping of electrons (Johnson *et al.* 1939). Experimental and theoretical studies of a more quantitative nature have been carried out by Randall and Wilkins (1945) and by Garlick and Wilkins (1945). Recently their hypotheses have been extended to explain the origin of the relatively large dielectric changes occurring in zinc sulphide phosphors during the emission of luminescence (Garlick and Gibson 1947).

The theoretical concept of a crystalline phosphor now generally accepted is that based on the "collective electron" model of Bloch which has been developed by various workers (see Mott and Gurney 1940). The allowed energies for electrons in such a phosphor consist of bands of energy states separated by forbidden energy bands as shown in figure 1 for the outermost electrons responsible for luminescence processes. Impurities, lattice defects or other perturbations of the ideal crystal lattice can give rise to discrete energy levels like those of an isolated atom and these may lie in the forbidden regions as shown. While impurity atoms, ions or larger complexes can function as luminescence emission centres the discrete levels of more minor disturbances can provide electron trapping levels just below the conduction levels of the system. These traps have been assumed to be those

responsible for the phenomena of phosphorescence and thermoluminescence. It is also generally assumed that in phosphors of the zinc sulphide type the electron traps are separate entities from the luminescence centres in contrast to the case of potassium chloride activated by thallium where the trapping states are within the thallium centres. Photoconductivity present in zinc sulphide phosphors supports this distinction.

The mechanism of electron trapping has been investigated in a relatively quantitative way by Randall and Wilkins (1945), and from their experiments they have postulated that, although the luminescence emission is due to recombination of excited electrons with empty luminescence centres, yet the experimental evidence favours assumption of a monomolecular process. As they have pointed out, this assumption is possible in explanations of phosphorescence and thermoluminescence, which are determined then only by the rate of escape of trapped electrons and their return to empty centres, if there is no retrapping of escaping electrons before their recombination with centres. Using the schematic model of figure 1 it is evident that the degree of retrapping will depend at any instant on the relative numbers of empty electron traps and empty luminescence centres and on their relative capture cross sections. Randall and Wilkins have shown that these capture cross sections are approximately equal and for normal phosphorescence conditions the numbers of empty electron traps and empty luminescence centres are about equal. Thus if electron traps and centres are independent entities the absence of retrapping cannot be explained by the model of figure 1.

Some of the conditions under which retrapping is very likely have been given by Garlick and Wilkins (1945) and depend on a high proportion of empty traps compared with the number of empty centres. These conditions occur at high temperatures where all traps are effectively empty, at the beginning of excitation of a previously unexcited phosphor and at very long decay times. The present studies are concerned with the investigation of the occurrence of retrapping under normal conditions as defined above and attempt to provide from experimental data a more correct theoretical model than previous ones for phosphors of the sulphide and silicate classes.

A note on the following studies has already been published (Garlick and Gibson 1946).

§ 2. THEORETICAL STUDIES

For the sake of completeness it is necessary to include and enlarge upon the theoretical developments of Randall and Wilkins (1945) following their assumption that the retrapping of electrons does not occur to any significant extent except for special conditions. If a phosphor contains only traps of one depth E and at any

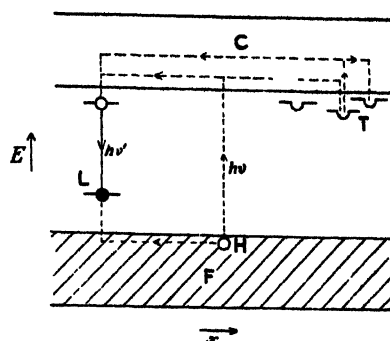


Figure 1. Schematic electron energy diagram for an impurity activated crystalline phosphor.

- C. Conduction energy band.
- F. Filled energy band.
- H. Positive hole.
- L. Luminescence centre.
- T. Electron traps.
- $h\nu$. Absorbed quantum.
- $h\nu'$. Emitted quantum.

time these contain n electrons, then the luminescence emission intensity I is determined by the rate at which electrons escape from traps, and the process is represented by the following equations:

$$I = -dn/dt = ns \exp(-E/kT), \quad \dots (1)$$

where $s \exp(-E/kT)$ is the probability per unit time that an electron escapes, s being a constant of the order of $10^{8 \pm 1} \text{ sec}^{-1}$ for most phosphors, E being the electron trap depth, k being Boltzmann's constant and T the absolute temperature. If the equations (1) are solved for a uniform rate of rise of temperature after excitation of the phosphor at a low temperature the variation of thermoluminescence intensity with temperature is given by the expression

$$I = n_0 s \exp(-E/kT) \cdot \exp \left\{ - \int_0^T \frac{s \exp(-E/kT)}{\beta} dT \right\}, \quad \dots (2)$$

where β is the rate of warming in deg/sec. Randall and Wilkins (1945) have discussed the effect of the warming rate on the thermoluminescence variation. The

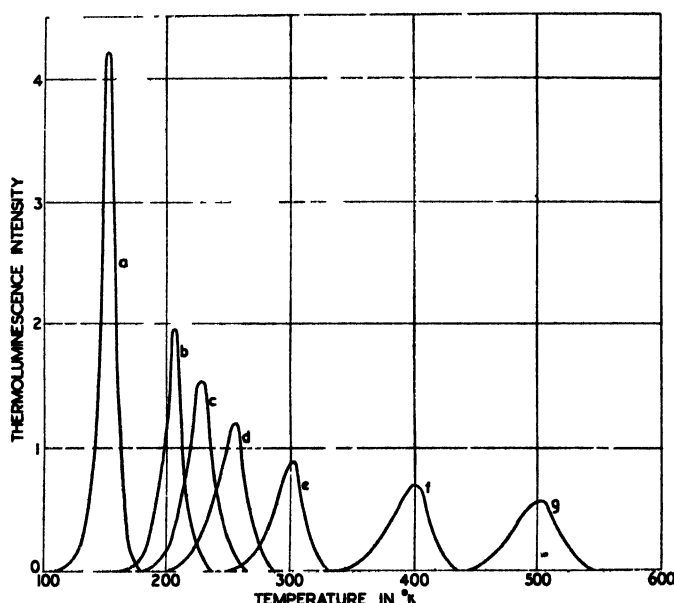


Figure 2(a). Variation of thermoluminescence with temperature for phosphors with single trap depths after excitation at 90°K . Curves calculated theoretically assuming retrapping to be absent. Warming rate $= 2.5^\circ/\text{sec}$.

- (a) $E = 0.3 \text{ ev.} : s = 10^9 \text{ sec}^{-1}$. (d) $E = 0.4 \text{ ev.} : s = 10^7 \text{ sec}^{-1}$. (f) $E = 0.8 \text{ ev.} : s = 10^9 \text{ sec}^{-1}$.
 (b) $E = 0.4 \text{ ev.} : s = 10^9 \text{ sec}^{-1}$. (e) $E = 0.6 \text{ ev.} : s = 10^9 \text{ sec}^{-1}$. (g) $E = 0.4 \text{ ev.} : s = 10^8 \text{ sec}^{-1}$.
 (c) $E = 0.4 \text{ ev.} : s = 10^8 \text{ sec}^{-1}$.

graphical form of equation (2) is shown in figure 2(a) for various values of s and E , n_0 and β remaining fixed ($\beta = 2.5^\circ/\text{sec}$). The chief characteristics of the curves of figure 2(a) for the variation of thermoluminescence with temperature when E is single valued are enumerated as follows:—

- (a) For given values of s and n_0 and β the temperature at which maximum emission occurs is proportional to the trap depth E .
 (b) For given values of E and n_0 the temperature of maximum emission varies with s and with β . For a change in s value from 10^9 sec^{-1} to 10^7 sec^{-1} when

$E=0.4$ ev. the position of the emission peak moves to higher temperatures by about 50°K . A hundred fold increase in the warming rate will cause the same shift.

(c) In all cases the area under the emission-temperature curve is proportional to the number of electrons initially trapped before warming begins, that is to n_0 . However, the actual shape of the emission curve is independent of n_0 which also means that the height of the curve at any point is proportional to n_0 .

(d) The initial rise of the thermoluminescence curve is exponential with temperature following the simple relation

$$I = n_0 s \exp(-E/kT). \quad \dots\dots(3)$$

This affords a means of determining E from experimental curves of the same form when s is not known.

By means of the above characteristics the experimental results from studies of phosphor specimens may be tested to see whether the phosphors have only one trap depth or a distribution of trap depths according to the above hypothesis.

The above theoretical developments rest on the assumption that retrapping is a negligible process in the phosphor. It is essential however to consider the modification of this theory when an electron escaping from a trap has the same probability of being retrapped as it has of recombining with an empty luminescence centre. There is evidence from previous studies that the two probabilities are about equal (Randall and Wilkins 1945, p. 405). Consider a phosphor containing a total number of electron traps N of which n are filled by electrons at any instant. Then there will be $(N-n)$ empty traps and n empty luminescence centres previously vacated by the trapped electrons (the number of electrons in the conduction band at any instant is assumed to be small compared with n during the phosphorescence and thermoluminescence). The probability that an escaping electron will recombine with an empty luminescence centre and not be retrapped is given by $n/\{(N-n)+n\}=n/N$ and the luminescence process will be given by

$$I = -dn/dt = (n^2/N)s \exp(-E/kT) \quad \dots\dots(4)$$

with the notation used above. These equations should be compared with the previous equations (2).

The decay of phosphorescence at a fixed temperature is given then by solution of equations (4) and is represented by the equation

$$I = n_0^2 s \exp(-E/kT) / N \{1 + (n_0/N)st \exp(-E/kT)\}^2. \quad \dots\dots(5)$$

The decay is hyperbolic in form showing that retrapping if present can produce a marked modification of the exponential decay given by equations (1) obtained by neglecting its effect. It will be seen from equation (5) that the form of the decay of phosphorescence will be dependent on n_0 and thus on the conditions of filling of the traps by excitation at the commencement of decay.

If equations (4) are solved to give the thermoluminescence variation with temperature when the phosphor is warmed at a uniform rate after excitation at a low temperature, then the resulting variation is given by the equation

$$I = n_0^2 s \exp(-E/kT) / N \left\{ 1 + \frac{n_0}{N} \int_0^T \frac{s \exp(-E/kT)}{\beta} dT \right\}^2. \quad \dots\dots(6)$$

This equation corresponds to equation (2) and its graphical forms for various values

of E and n_0 when s and β remain fixed are given in figure 2(b). The main characteristics of thermoluminescence-temperature variation when retrapping occurs are enumerated as follows :—

(a) For fixed values of s , n_0 and β the temperature at which the emission reaches its maximum is proportional to the electron trap depth E . For saturation of the traps ($n_0 = N$) and the same values of s the equations (2) and (6) give maxima at the same temperature.

(b) For given values of E and n_0 the temperature position of the emission maximum varies with s and β in much the same way as for the case when retrapping is absent.

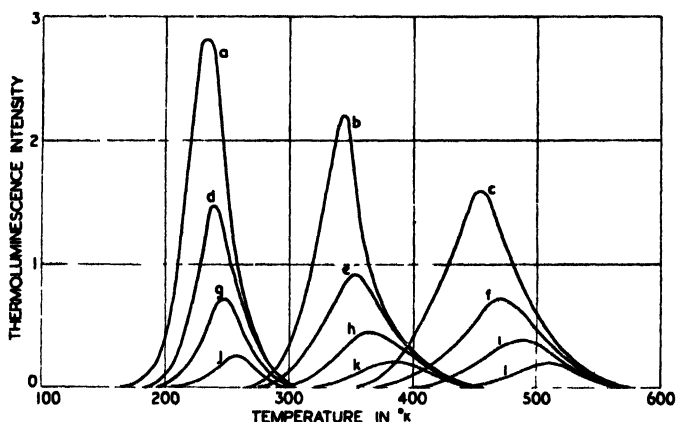


Figure 2(b). Variation of thermoluminescence with temperature for phosphors with single trap depths after excitation at 90 °K. Curves calculated theoretically assuming retrapping to be present. Warming rate = 2.5 °K/sec. $s = 10^4 \text{ sec}^{-1}$, N = total number of traps of depth E ; n_0 = number filled.

- | | | |
|---|--|--|
| (a) $E = 0.4 \text{ ev.} : n_0 = N.$ | (e) $E = 0.6 \text{ ev.} : n_0 = 0.5N.$ | (i) $E = 0.8 \text{ ev.} : n_0 = 0.25N.$ |
| (b) $E = 0.6 \text{ ev.} : n_0 = N.$ | (f) $E = 0.8 \text{ ev.} : n_0 = 0.5N.$ | (j) $E = 0.4 \text{ ev.} : n_0 = 0.1N.$ |
| (c) $E = 0.8 \text{ ev.} : n_0 = N.$ | (g) $E = 0.4 \text{ ev.} : n_0 = 0.25N.$ | (k) $E = 0.6 \text{ ev.} : n_0 = 0.1N.$ |
| (d) $E = 0.4 \text{ ev.} : n_0 = 0.5N.$ | (h) $E = 0.6 \text{ ev.} : n_0 = 0.25N.$ | (l) $E = 0.8 \text{ ev.} : n_0 = 0.1N.$ |

(c) In all cases the area under the emission-temperature curve is proportional to the number of electrons (n_0) initially trapped. However, the actual shape of the curve is dependent on n_0 as shown by equation (6). This is essentially different from the case when retrapping is negligible.

(d) The initial rise of luminescence before peak emission is reached follows the simple relation

$$I = (n_0^2/N)s \exp(-E/kT). \quad \dots\dots (7)$$

Comparing this equation with equation (3) it is seen that the power of the n_0 term is different in the two equations, the exponential term being identical. This provides a further means of experimental correlation with theory.

The above characteristics obtained by the two different assumptions of retrapping and no retrapping show the marked effect which retrapping may have on the phosphor characteristics. For most phosphors which have complex electron trap distributions the above treatment must be extended to traps of many depths. This is relatively simple if retrapping is neglected as shown by Randall and Wilkins for particular trap distributions. For example if the number of traps decreases

exponentially with depth (i.e. $N \propto e^{-\alpha E}$) then the phosphorescence decay for initial saturation of the traps involved is of the form $I \propto t^{-(\alpha kT + 1)}$ where t is the decay time. The trap distributions may be obtained by means of thermoluminescence experiments and in the above case the theory verified by correlation of the results with phosphorescence measurements at different fixed temperatures, since the index of t should vary linearly with the temperature. This particular case shows that the hyperbolic decays characteristic of zinc sulphide phosphors can be explained by the sum of exponential terms whose constants are known from experiments. The derivation of phosphorescence characteristics for phosphors with complex distributions when retrapping is present becomes difficult as electrons escaping from traps of one depth may be retrapped in traps of different depths. Solutions have not yet been obtained for these conditions.

It is necessary to consider the way in which the trapping processes are affected by the conditions of excitation of the phosphor both with and without the presence of retrapping of escaping electrons. As a simple case that of steady excitation of a phosphor at fixed temperatures will be considered.

(i) *The filling of traps by constant excitation with no retrapping*

The considerations below make a fundamental change in the theoretical model of figure 1. They assume that each luminescence centre has an electron which on excitation is raised into electron traps which are in the immediate neighbourhood of the centre or are metastable trapping levels within the centre. Then if ultra-violet light of intensity J excites aJ electrons per second and in equilibrium the number of electrons trapped is n_0 in the N available traps of depth E then the equilibrium is represented by

$$dn_0/dt = 0 = abJ(N - n_0) - n_0s \exp(-E/kT) \quad \dots\dots(9)$$

where b is the probability of electron capture by a trap. This gives the relation between n_0 and N for different intensities of ultra-violet radiation, temperatures and trap depths as

$$n_0 = N / \{1 + (s/abJ) \exp(-E/kT)\}. \quad \dots\dots(10)$$

This equation may be tested experimentally by both thermoluminescence and phosphorescence experiments. It may be extended to apply to complex trap distributions as has been done in the theoretical treatment of dielectric changes due to electron traps in phosphors (Garlick and Gibson 1947). For the purposes of the present experimental studies equation (10) is rearranged to show the simple relation between n_0 and J for fixed trap depth and temperature

$$(N/n_0) - 1 = (s/abJ) \exp(-E/kT). \quad \dots\dots(11)$$

(ii) *The filling of electron traps when retrapping is present*

With the above notation and assuming retrapping to be present as in the derivation of equations (4), (5) and (6), the filling of the N available traps by constant excitation J at a fixed temperature may be determined. If aJ electrons are raised into the conduction band per second by the excitation and in equilibrium m_0 electrons are in the band, n_0 being in traps, then $(n_0 + m_0)$ centres are empty and the equilibrium equations are

$$\begin{aligned} dm_0/dt &= 0 = aJ - \beta m_0(m_0 + n_0), \\ dn_0/dt &= 0 = bm_0(N - n_0) - n_0s \exp(-E/kT), \end{aligned}$$

where b is the probability of capture of an electron by an empty trap and β the

probability of electron recombination with an empty luminescence centre. Solution of these equations shows that below saturation level for the traps (i.e. $N > \dot{n}$) both m_0 and n_0 are root functions of the excitation intensity J . Thus when retrapping is present the simple relation between n_0 and J as in equations (10) and (11) is not found and the examination of trap filling thus provides a means of testing the above relations for validity.

The above theoretical developments have been compared with the results of experimental studies on typical phosphors as described below and these comparisons provide evidence for the relative importance of retrapping.

§ 3. EXPERIMENTAL STUDIES AND RESULTS

(i) *Experimental methods*

The experiments consisted essentially of measurements of thermoluminescence and phosphorescence characteristics of selected specimens under many different but known conditions of excitation and temperature. For all these experiments the phosphor was mounted *in vacuo* in a thin layer on a rhodium-plated copper surface forming part of a metal Dewar vessel described in a previous paper (Garlick and Wilkins 1945). The specimens could be excited by monochromatic radiation of wavelengths 2537 Å. or 3653 Å. by means of a mercury arc source. The methods of measuring thermoluminescence curves and phosphorescence curves have also been previously described (Randall and Wilkins 1945). The rate of warming in all thermoluminescence experiments was kept fixed at 2.5°/sec.

(ii) *Experimental results and theoretical correlations*

For the experimental investigations three phosphor specimens have been chosen with narrow trap distributions giving a peak in their saturated thermoluminescence curves which are shown in figures 3 (a), 3 (b) and 3 (c). The specimen corresponding to figure 3 (a) is a zinc sulphide phosphor activated by copper and excited by 3653 Å. radiation. It has two peaks in its thermoluminescence curve but in experimental studies use was made of the peak due to deeper traps only. Figure 3 (b) is the thermoluminescence curve for a zinc silicate phosphor activated by manganese and excited by 2537 Å. radiation having a single peak in its emission curve. In both figures 3 (a) and 3 (b) the theoretical curves obtained by substitution of suitable values of s and E in equation (6) are shown by broken curves. The experimental curves show remarkably good agreement with the calculated curves for which retrapping has been assumed. However, on the basis of "no trapping" these phosphors would have a distribution of trap depths. Both specimens are typical of their class and are photoconductors. The third specimen studied is a strontium silicate phosphor with europium impurity as activator. Its thermoluminescence curve has one main peak as shown in figure 3 (c) and this represents a single trap depth as shown by the agreement with the theoretical curve, the broken curve of figure 3 (c), derived from equation (2) assuming no retrapping. This phosphor is excited by ultra-violet light of both wavelengths used but thermoluminescence and the corresponding phosphorescence of long duration only occurs after excitation by 2537 Å. radiation. The unique characteristics of the luminescence of this phosphor are described and discussed later. The following experiments attempt to correlate the characteristics of each phosphor with one or other of the theoretical hypotheses. By this means the extent to which retrapping occurs may be defined.

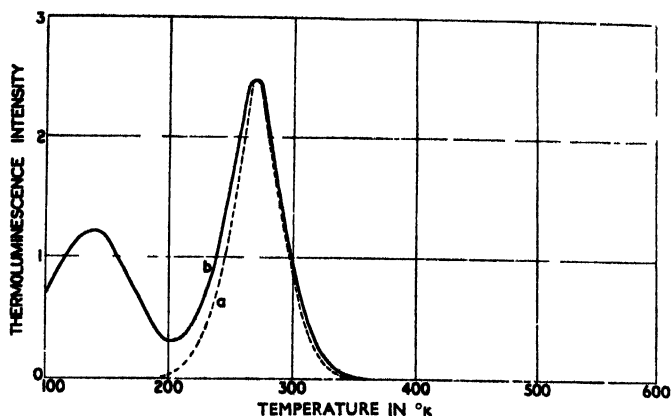


Figure 3 (a). ZnS—Cu phosphor. Thermoluminescence characteristics after excitation at 90° K.

(a) Theoretical curve assuming retrapping to occur. (b) Experimental curve.

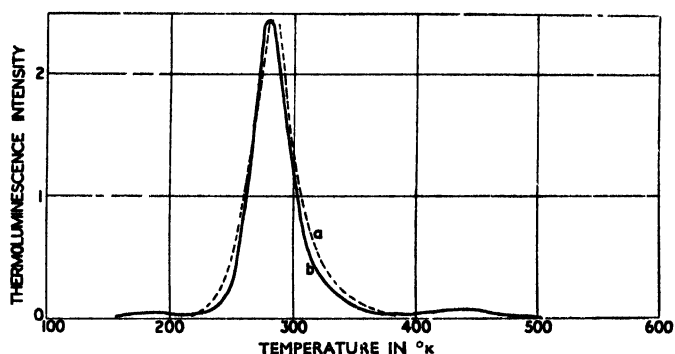


Figure 3 (b). Zn_3SiO_4 —Mn phosphor. Thermoluminescence characteristics after excitation at 90° K.

(a) Theoretical curve assuming retrapping to occur. (b) Experimental curve.

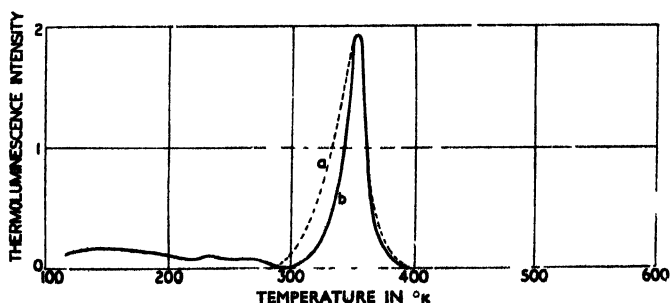


Figure 3 (c). SrSiO_3 —Eu phosphor. Thermoluminescence characteristics after excitation at 90° K.

(a) Theoretical curve assuming retrapping to be absent. (b) Experimental curve.

§ 4. THE DECAY OF PHOSPHORESCENCE AT FIXED TEMPERATURE

(i) *Copper-activated zinc sulphide*

The decay of phosphorescence of the zinc sulphide phosphor has been measured at several different fixed temperatures. The decay as measured at room temperature is shown in figure 4, curve *a*, the coordinates of the graph being logarithmic to show the hyperbolic form of the decay, at longer times. At room temperature for these longer times the intensity of phosphorescence is inversely proportional to the 2.2 power of the decay time. This power is found to vary linearly with temperature as shown by the measured values at different temperatures given in table 1.

Table 1. Power law indices for the phosphorescence decay of ZnS-Cu phosphor

Temperature ($^{\circ}$ K.)	265	292	318	355
Power law index (αkT^{-1})	2.07	2.2	2.32	2.5
α (ev $^{-1}$)	47	47.7	48	49
α from thermoluminescence measurements (ev $^{-1}$)	49	49	49	49

These experimental results indicate that the phosphor does not exhibit the phosphorescence characteristics predicted by equation (5) although its thermoluminescence curve may be interpreted by equation (6) since the final asymptotic form of the decay varies with temperature. However, if the thermoluminescence curve and phosphorescence decay curves are interpreted according to the hypothesis which neglects retrapping, then the linear variation of the power law index is expected from equation (8), and from the decay curves α of the equation may be obtained. The constant α may also be obtained from the exponential form of the thermoluminescence curve at its higher temperature end given in figure 3 (*a*). Values of α derived from the decay curves are given in the third line of table 1 and the value from the curve of figure 3 (*a*) in the fourth line of the table. In the latter case the electron trap constant s had the value 10^8 sec^{-1} found by other experiments as previously described (Randall and Wilkins 1945, Garlick and Gibson 1947). Good correlation is obtained between the differently derived values of α . Thus experimental evidence from decay measurements shows that the deeper traps in this phosphor, while giving rise to a relatively narrow thermoluminescence-temperature curve, have a distributed range of depths and that agreement of experimental results with theory necessitates the assumption that retrapping is negligible.

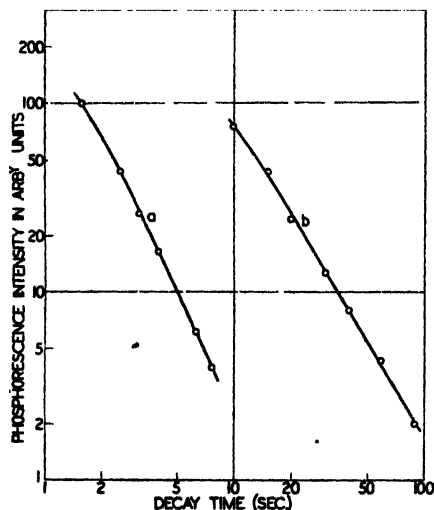


Figure 4. Decay of phosphorescence with time for zinc sulphide and zinc silicate phosphors at room temperature (291° K.).

- (*a*) Copper activated zinc sulphide specimen.
(*b*) Manganese activated zinc silicate specimen.

(ii) *Manganese-activated zinc silicate*

The decay of phosphorescence of this specimen at room temperature is shown by curve (b) of figure 4. It is of the similar hyperbolic form to curve (a), at longer decay times its power law index being 1.7. By use of equation (8) this gives the trap distribution constant α as 28.0 ev^{-1} . Derivation from the thermoluminescence curve of figure 3 (b) gives α as 25 ev^{-1} . Again agreement with theory is good when retrapping is neglected. The value of the power law index is found to increase linearly with temperature as in the case of the zinc sulphide phosphor. These facts provide the data supporting the theoretical hypothesis that retrapping is not important.

(iii) *The phosphorescence of europium-activated strontium silicate*

This phosphor may be excited to give fluorescence and phosphorescence by radiation of either 3653 Å. or 2537 Å. However, there is no long duration phosphorescence or any thermoluminescence when the longer wavelength radiation is

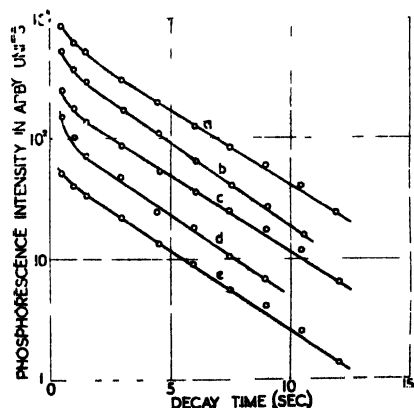


Figure 5 (a). Phosphorescence decay of $\text{SrSiO}_3\text{—Eu}$ phosphor at various temperatures after excitation by 3650 Å. radiation.

- (a) 476° K. (d) 192° K.
 (b) 398° K. (e) 90° K.
 (c) 294° K.

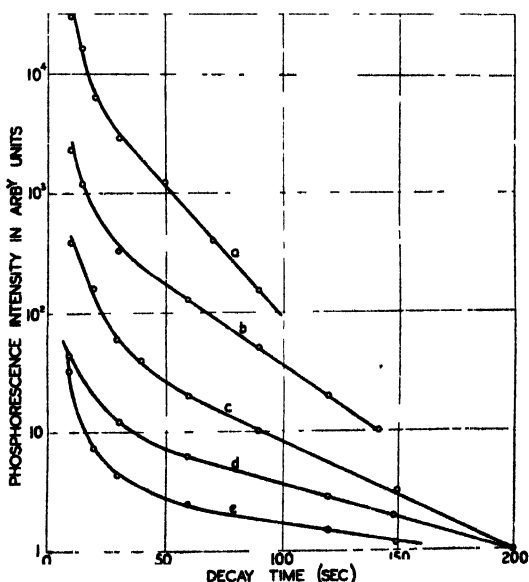


Figure 5 (b). Phosphorescence decay of $\text{SrSiO}_3\text{—Eu}$ phosphor at various temperatures after excitation by 2537 Å. radiation.

- (a) 387° K. (d) 340° K.
 (b) 373° K. (e) 320° K.
 (c) 358° K.

used. The phosphorescence decay with time has been measured after excitation by each of these wavelengths of ultra-violet radiation. Figure 5 shows the results obtained. Figure 5 (a) shows the form of the phosphorescence decay after excitation by radiation of 3635 Å. wavelength at different temperatures from 90° K. to 476° K. The initial decay is rapid, occupying a small fraction of a second. The subsequent decay has an exponential form, the decay rate being independent of temperature. The half-value period of the decay at different temperatures from

90° K. to 476 K. is included in the figure, its mean experimental value being 2.25 sec. Since this decay rate is independent of temperature it cannot be due to thermally metastable electron energy levels. This is also supported by the absence of thermoluminescence and photoconductivity when the 3650 Å. is used. Thus the phosphorescence must be due to a forbidden optical transition within the europium luminescence centre, the life time of the state being abnormally large. Similar decay characteristics have been observed in other studies of the phosphorescence of adsorbed dyestuffs made according to the method of Travnicek (1937).

When radiation of wavelength 2537 Å. is used to excite the specimen the latter exhibits both photoconductivity and thermoluminescence. Its phosphorescence decay at various temperatures is given in figure 5(b). The initial decay is due to the same two components which are present when 3653 Å. irradiation is used, but the decay includes also phosphorescence of long duration decaying exponentially with time over the range measured but having a decay rate markedly dependent on temperature. This decay must be due to electron traps with a single depth or at the most a very narrow spread of trap depths as indicated by the thermoluminescence curve of figure 3(c). The absence of a hyperbolic decay due to these electron traps provides evidence that the hypothesis of Randall and Wilkins applies in this case. The phosphorescence decay is thus of the form:—

$$I = c \exp(-st e^{-E/kT}). \quad \dots\dots (14)$$

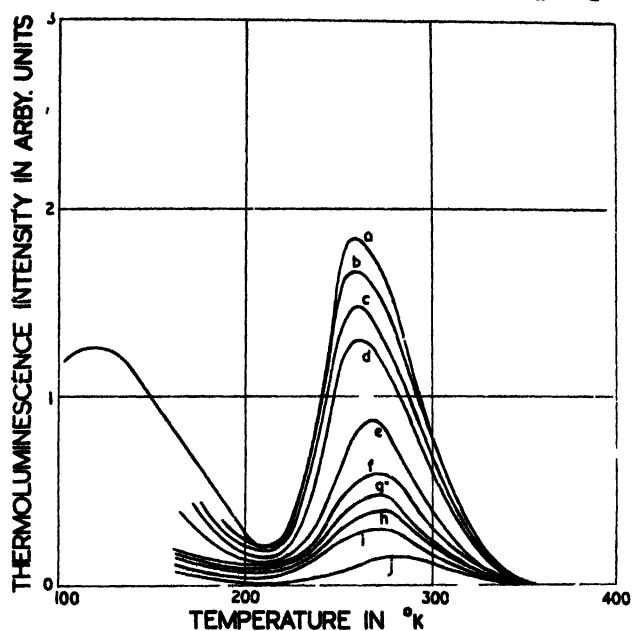
From the change of decay rate the value of E has been calculated and is found to be about 0.32 eV. This value is not in agreement with those obtained from thermoluminescence measurements and is discussed in the section below.

§ 5. STUDIES OF THERMOLUMINESCENCE CHARACTERISTICS

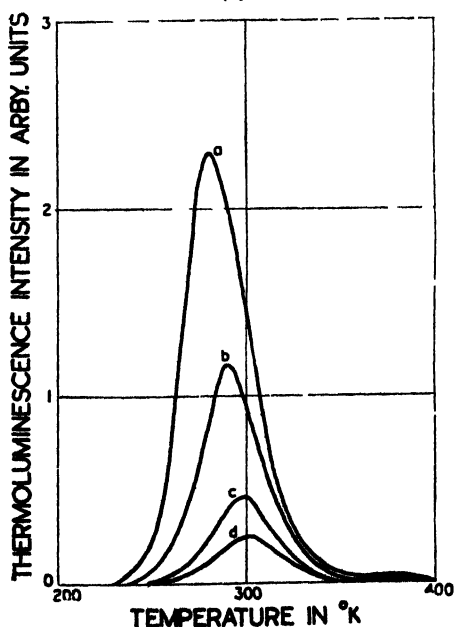
The measurement of the variation of the thermoluminescence characteristics of a phosphor with different intensities of initial excitation provides a means of determining how the electron traps are filled for the different excitation conditions. Each phosphor specimen is therefore excited at liquid air temperature with various intensities of ultra-violet light of suitable wavelength for a sufficiently long time so that equilibrium conditions are reached. Then after excitation is removed the thermoluminescence-temperature curves are measured. These curves are given in figures 6(a), (b) and (c). The area under each curve represents the number of electrons initially trapped, denoted by n_0 , under the equilibrium excitation conditions at 90° K. It will be seen from these curves that the temperature at which maximum emission occurs does not alter appreciably as the excitation intensity is varied.

In order to compare these experimental results with the theoretical predictions represented by equations (11) and (14) the area under each curve is taken as n_0 , the number of electrons initially trapped, and the area under the curve for very high intensities as the saturation value N . Then for each specimen the experimentally derived values of $((N/n_0) - 1)$ are plotted against the inverse of the excitation intensity J . These plots are given in figures 7(a), (b) and (c) for the three specimens respectively. From inspection it is found that the experimental results follow the relations given by equation (11), i.e. retrapping assumed to be absent.

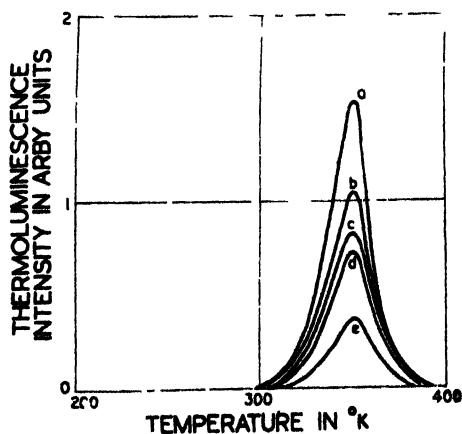
A further correlation of experiment with theory can be made by considering the portion of the thermoluminescence curve at some temperature lower than that at which peak emission occurs. As shown by the equations (3) and (7) the shape



- (a) Zinc sulphide—Copper activated phosphor.
- | | | |
|----------------|---------------|----------------|
| (a) $J=1740$. | (e) $J=3$. | (h) $J=0.97$. |
| (b) $J=69$. | (f) $J=1.6$. | (i) $J=0.93$. |
| (c) $J=27$. | (g) $J=1.0$. | (j) $J=0.53$. |
| (d) $J=9$. | | |



- (b) Zinc silicate—Manganese-activated phosphor.
- | | |
|--------------|-------------|
| (a) $J=27$. | (c) $J=3$. |
| (d) $J=9$. | (d) $J=1$. |



- (c) Strontium sulphide—Manganese-activated phosphor.
- | | |
|---------------|-------------|
| (a) $J=100$. | (d) $J=3$. |
| (b) $J=27$. | (e) $J=1$. |
| (c) $J=9$. | |

Figure 6. Variation of thermoluminescence of sulphide and silicate phosphors with intensity of excitation (J) at 90°K .

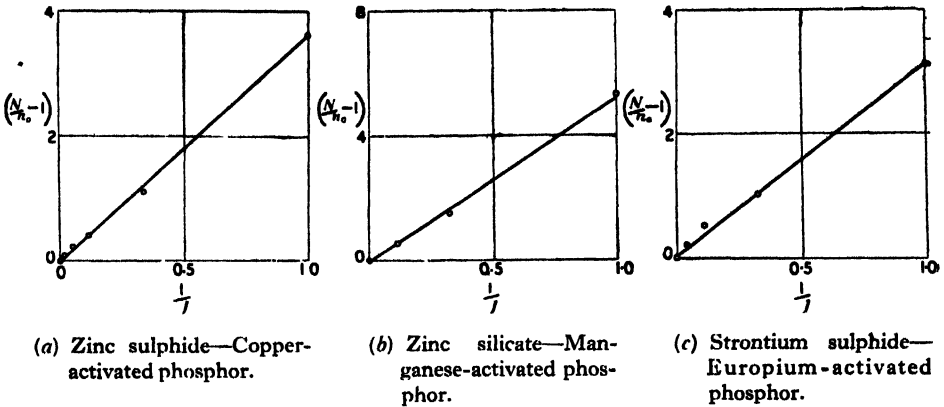


Figure 7. Variation in the number of trapped electrons (n_0) in the number of available traps (N) with excitation intensity (J) at 90°K .

of this portion of any curve will depend upon the degree of retrapping present. From this shape the value of trap depth may also be obtained for each specimen. Figures 8(a), (b) and (c) show how the height of the initial portion of the curve at some selected fixed temperature varies with the number of electrons initially trapped, as given by the total area under the curve. Results for each of the three specimens show that equation (3) is adhered to. Calculations of the trap depth from the curve shape have been made for the strontium silicate specimen where the depth is almost single-valued. The depth thus obtained is 0.68 eV. By using this value and the temperature of peak emission the value of the constants may be found. This is calculated to be 10^9 sec^{-1} which is of the usual order for silicate and sulphide phosphors (Randall and Wilkins 1945, Garlick and Gibson 1947). Again by using this value of the constant the value of E may be calculated from the phosphorescence decay curves of figure 5(b) since the decay rate is given by $\exp(-E/kT)$. The values thus obtained are collected in table 2.

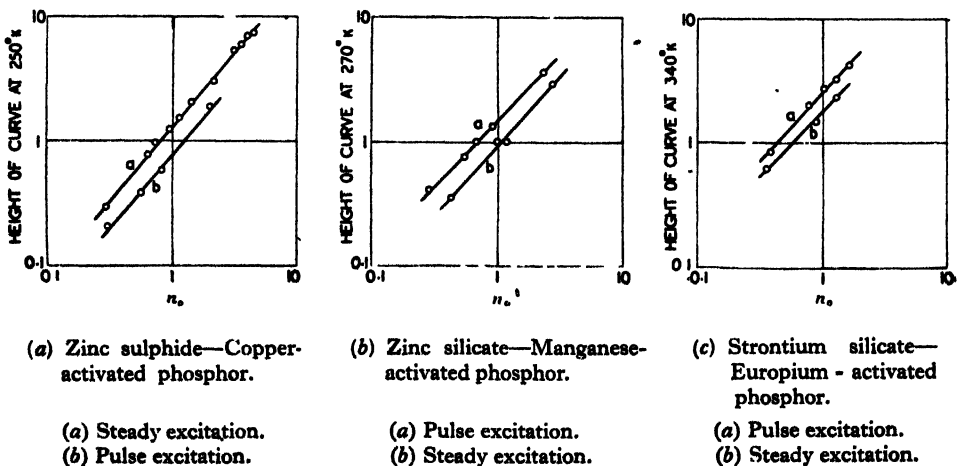


Figure 8. Variation in height of initial portion of thermoluminescence curve with the number of electrons (n_0) initially trapped by excitation at 90°K .

Table 2. Values of electron trap depth for specimen of europium-activated strontium silicate obtained from phosphorescence characteristics with constant $s = 10^9 \text{ sec}^{-1}$.

Temperature ($^{\circ} \text{K.}$)	320	340	358	373	387
Decay rate $s \exp (-E/kT)$	0.00935	0.013	0.021	0.031	0.05
Electron trap depth E (ev.)	0.74	0.74	0.75	0.77	0.77

As shown by the phosphorescence curves the values of E given in table 1, while agreeing approximately with that obtained from the thermoluminescence curve, do not correlate with the value obtained from the change of decay rate with temperature. This may be due to a small spread of E values for this specimen similar to that found by Randall and Wilkins (1945) for potassium chloride activated by thallium. Otherwise if E is considered single valued the value of the constant s must be of the order of 10^4 sec^{-1} according to decay rate changes, which does not agree with the value obtained from the thermoluminescence measurements.

(ii) *Changes in thermoluminescence characteristics during phosphorescence decay and after pulse excitation*

The following experiments provide further data to distinguish between re-trapping and no-retrapping processes in the phosphor specimens and follow the method used in earlier studies (Garlick and Wilkins 1948). Each specimen is excited at liquid oxygen temperature by a pulse of ultra-violet light whose duration can be varied. After such excitation the variation of the thermoluminescence with temperature is measured. The curves obtained are shown in figures 9(a), (c) and (e) for the three phosphor specimens respectively. The specimens are also excited at room temperature (291°K.) and allowed to phosphoresce after removal of excitation for various selected times after which periods the specimen is suddenly cooled and then warmed at the uniform rate of $2.5^{\circ}/\text{sec.}$, its thermoluminescence

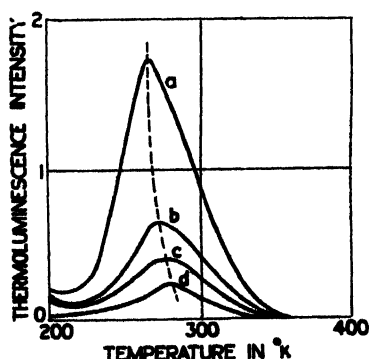


Figure 9(a). ZnS—Cu phosphor. Variation of thermoluminescence with duration of pulse excitation at 90°K.

- (a) Steady excitation.
- (b) Medium length pulse.
- (c) Short pulse.
- (d) Very short pulse.

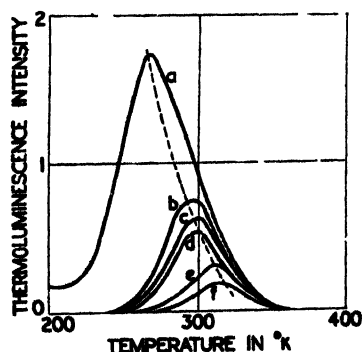


Figure 9(b). ZnS—Cu Phosphor. Variation of thermoluminescence after different times of phosphorescence decay at 291°K.

- (a) Excitation at 90°K.
- (b) Decay for 1 sec. at 291°K.
- (c) Decay for 2 sec. at 291°K.
- (d) Decay for 3 sec. at 291°K.
- (e) Decay for 10 sec. at 291°K.
- (f) Decay for 20 sec. at 291°K.

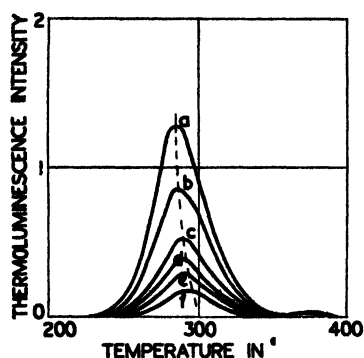


Figure 9 (c). Zn_2SiO_4 —Mn phosphor. Variation of thermoluminescence with duration of pulse excitation at 90 K.

- (a) Steady excitation.
- (b) Long pulse.
- (c) Medium pulse.
- (d) Short pulse.
- (e) Short pulse.
- (f) Very short pulse.

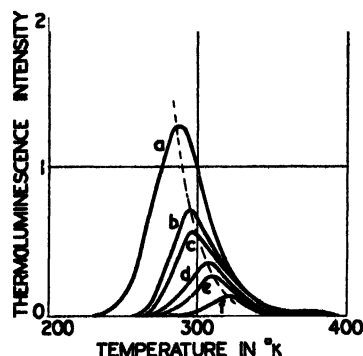


Figure 9 (d). Zn_2SiO_4 —Mn phosphor. Variation of thermoluminescence after different times of phosphorescence decay at 291° K.

- (a) Excitation at 90° K.
- (b) Decay for 1 sec. at 291° K.
- (c) Decay for 4 sec. at 291° K.
- (d) Decay for 10 sec. at 291° K.
- (e) Decay for 20 sec. at 291° K.
- (f) Decay for 60 sec. at 291° K.

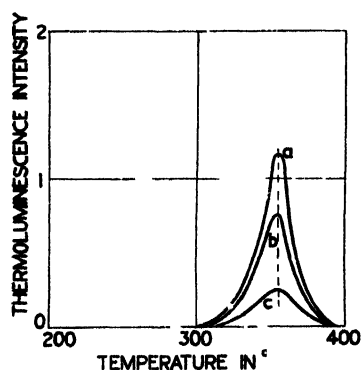


Figure 9 (e). SrSiO_3 —Eu phosphor. Variation of thermoluminescence with duration of pulse excitation at 90 K.

- (a) Steady excitation.
- (b) Medium pulse.
- (c) Short pulse.

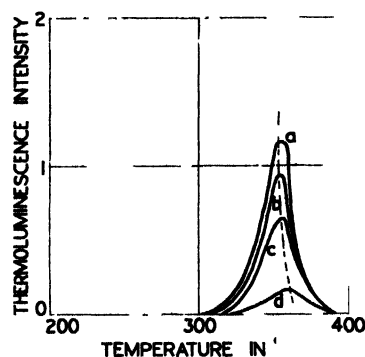


Figure 9 (f). SrSiO_3 —Eu phosphor. Variation of thermoluminescence after different times of phosphorescence decay at 291° K.

- (a) Excitation at 90° K.
- (b) Decay for 30 min. at 291° K.
- (c) Decay for 1.5 hr. at 291° K.
- (d) Decay for 16 hr. at 291° K.

temperature variation being measured during warming. The curves so obtained are given in figures 9(b), (d) and (f) and represent the distribution of electrons in traps at different stages of phosphorescence decay. Cross comparison between curves 9(a), and (b), (9c) and (d) and (9e) and (f) respectively shows that change in duration of pulse excitation affects the number of trapped electrons but the shape of the thermoluminescence curves and their maxima positions are not markedly altered. However, phosphorescence decay does cause a shift of the maximum position of these curves at higher temperatures as the decay proceeds for the specimens of zinc sulphide and zinc silicate. The strontium silicate phosphor characteristics in curves 9(e) and 9(f) show no change in the maximum position

with the number of trapped electrons either during phosphorescence or due to pulse excitation.

These results confirm the hypothesis that retrapping is a negligible process in these phosphors. As stated in the theoretical section above the thermoluminescence curve shape for a single trap depth when retrapping is present depends on the number of trapped electrons (figure 2(b)) but is independent of the process of filling or emptying, whether by pulse excitation or phosphorescence decay. Thus for a given area under the curve the shape should be fixed. As shown by comparison of curves (a) and (b), (c) and (d), (e) and (f) respectively of figure 9 the experimental results do not agree with this prediction but are in agreement with the supposition that a distribution of trap depths is present and that retrapping is negligible. This means that pulse excitation of the previously unexcited phosphor will fill traps of all different depths with equal probabilities, the number filled thus depending only on the number available. Thus the shape of the thermoluminescence curve is little affected. However, when phosphor-decay proceeds the shallower traps are emptied leaving only the deeper ones filled with the consequence observed in the curves of figure 9(b), (d) and (f), that is, a temperature shift of the emission maxima.

§ 6. CONCLUSION

The experimental results described above have been obtained from studies of three typical phosphor specimens of three different classes. In each case the specimen has been selected because it affords suitable thermoluminescence and phosphorescence properties to determine how far retrapping of electrons escaping from electron traps was a major part of the mechanism. The evidence presented has shown that the characteristics measured are in agreement with the simple theoretical treatment originally due to Randall and Wilkins which assumes retrapping to be negligible. This could be understood in terms of the model of figure 1 if, as in the case of the thallium activated alkali halide phosphors, the electron traps are within the luminescence centres to which the phosphorescence processes would then be confined. However, the specimens all exhibit photoconductivity which is associated with the thermoluminescence and phosphorescence processes and therefore electrons must leave luminescence centres during the process. This means that the luminescence process involves a recombination of electrons with luminescence centres after escaping from electron traps which are not in the luminescence centres themselves. Previous experiments due to Randall and Wilkins (1945) have given estimates of the capture cross sections of empty luminescence centres and electron traps which appear to be of the same order. Thus retrapping of escaping electrons from traps will be an integral part of the mechanism if traps are quite independent of luminescence centres. It is therefore more probable that the impurity centres while not containing the trapping energy states cause perturbations in their environment (i.e. in the surrounding local lattice) which give rise to the trapping states. If this is so then the absence of retrapping of electrons may be more readily understood since electrons escaping from traps adjacent to empty luminescence centres will have a higher probability of recombining with the centres than of moving to other empty traps and being recaptured. It is not possible to provide a detailed treatment of this assumption as no studies of the complex potential configurations around impurity centres in a matrix lattice have been made. However, the postulate of association between traps and

luminescence centres is supported by other and independent studies of the effect of one or more different impurities and also of the preparation conditions (firing temperature, fluxes etc.) on the electron trap distributions of phosphors (Garlick 1947). These experimental investigations show that introduction of specific impurities gives rise to particular electron trap configurations. Other studies of the dielectric changes in zinc sulphide and other phosphors show good agreement between theory and experiment when retrapping is neglected (Garlick and Gibson 1947).

To summarize, the experimental and theoretical developments described above show that the simple electron trap theory due to Randall and Wilkins (1945) which neglects retrapping is in better agreement with experiment than that in which retrapping is assumed to be present. This evidence, combined with that from other investigations leads to a new concept of the origin and spatial location of electron traps in phosphors. They appear to be formed by the perturbations due to the impurity activator in the matrix crystal lattice immediately surrounding the centre.

ACKNOWLEDGMENTS

Our thanks are due to Professor M. I. Oliphant, F.R.S., for the provision of facilities for this work and to the Ministry of Supply (D.C.D.) under whose research projects it has been carried out.

REFERENCES

- GARLICK, G. F. J., 1947, *Monograph of Cornell Conference on Luminescence* (New York : John Wiley).
- GARLICK, G. F. J., and GIBSON, A. F., 1946, *Nature, Lond.*, **158**, 704 ; 1947, *Proc. Roy. Soc. A*, **188**, 485.
- GARLICK, G. F. J., and WILKINS, M. H. F., 1945, *Proc. Roy. Soc. A*, **184**, 408.
- DE GROOT, W., 1939, *Physica*, **6**, 275.
- JOHNSON, R. P., 1939, *J. Opt. Soc. Amer.*, **29**, 387.
- MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals* (Oxford : University Press).
- RANDALL, J. T., and WILKINS, M. H. F., 1945, *Proc. Roy. Soc. A*, **184**, 366.
- TRAVNICÛK, M., 1937, *Ann. Phys., Lpz.*, **30**, 224.

The Moving Coil Galvanometer as a Circuit Element

By N. F. ASTBURY

G.K.N. Research Laboratories, Wolverhampton

MS. received 2 September 1947 ; read 6 February 1948

ABSTRACT. An account is given of a simplified calculus applicable to a moving-coil galvanometer in which use is made of the concept of motional impedance. The fundamental galvanometer equations are presented in a novel and very simple form. Examples of the use of the formulae are given, notably with reference to the ultimate "noise level" of a galvanometer.

§ 1. INTRODUCTION

THE moving-coil galvanometer, one of the oldest and most familiar of our laboratory instruments, has recently found considerable application in various systems for direct-current amplification, amongst which may be noted those described by D. C. Gall (1942, 1945). It is hoped to publish

shortly descriptions of other systems developed by the present writer, but in the meantime it may be noted that the problem of D.C. amplification is one of very great technical importance and one which is most admirably met by the use of a moving-coil galvanometer in a preamplifying stage. In the course of development work on problems of this type it occurred to the writer that the mathematics of a moving-coil galvanometer could be considerably simplified for many purposes of circuit analysis, and it is the purpose of this paper to show how the behaviour of a galvanometer as a circuit element can be succinctly expressed in terms of its motional impedance, which can in turn be expressed in terms of experimentally determined constants.

While the concept of motional impedance applied to galvanometers is by no means new, it is believed that the present treatment is novel. Butterworth (1912) showed that a galvanometer can be represented by a parallel combination of a pure inductance, a capacitance and resistance placed in series with a resistance equal to that of the instrument coil, and that in this equivalent circuit the current in the inductance branch is homologous with the angular deflection of the coil. Kennelly showed that the impedance circle diagram was applicable to a galvanometer. A note on these points is given in the *Dictionary of Applied Physics* (Vol. II, p. 976).

It is proposed in the present paper to put forward the motional impedance concept and to illustrate it by applications to various circuit problems.

§ 2. PRÉCIS OF STANDARD RESULTS

It is convenient to start with a précis of well-known results for the galvanometer.

Let α , β and γ be respectively the moment of inertia of the moving system, the mechanical damping factor and the torque control constant of the galvanometer, and let k be the product of the total area of the coil and the control field, i.e. the "area-turns-field" constant. Then, if the galvanometer form part of a closed circuit of total resistance R (and of zero reactance) in which an E.M.F. e is acting, the current i and the angular deflection θ of the moving system are related to e by the torque and E.M.F. equations

$$(\alpha D^2 + \beta D + \gamma)\theta = ki, \quad \dots\dots(2.1)$$

$$Ri + kD\theta = e, \quad \dots\dots(2.2)$$

where D stands for d/dt .

Eliminating i , we have

$$\{\alpha D^2 + (\beta + k^2/R)D + \gamma\}\theta = ke/R, \quad \dots\dots(2.3)$$

from which we find S , the critical resistance, as

$$S \simeq k^2/2\alpha\omega_0 \quad \dots\dots(2.4)$$

where ω_0 is the free frequency of the instrument and is given by

$$\omega_0 = (\gamma/\alpha)^{1/2}(1 - \beta^2/4\alpha\gamma)^{1/2} \simeq (\gamma/\alpha)^{1/2}. \quad \dots\dots(2.5)$$

The logarithmic decrement, λ , is, if $T = 2\pi/\omega$,

$$\lambda = (\beta + k^2/2R)(T/2\alpha) \quad \dots\dots(2.6)$$

and

$$\lambda_0 = \beta\pi/\alpha\omega_0, \quad \dots\dots(2.7)$$

where λ_0 is the value of λ for the open circuit condition. We may define the damping factor, δ , by

$$\delta = \lambda_0/\pi = \beta/\alpha\omega_0. \quad \dots\dots(2.8)$$

Finally, we may define the direct current sensitivity, μ , by

$$\mu = k/\gamma \quad \dots\dots (2.9)$$

and then express all our coefficients in terms of the experimentally observable quantities S , ω_0 and μ , obtaining

$$\alpha = 2S'\omega_0^3\mu^2, \quad \dots\dots (2.10)$$

$$\beta = 2S\lambda_0/\pi\omega_0^2\mu^2, \quad \dots\dots (2.11)$$

$$= 2S\delta_1\omega_0^3\mu^2, \quad \dots\dots (2.12)$$

$$\gamma = 2S_1\mu^2\omega_0, \quad \dots\dots (2.13)$$

$$k = 2S_1\mu\omega_0. \quad \dots\dots (2.14)$$

These results can be used to express Butterworth's (1912) formulae in simple forms. If L , C and P are respectively the inductance, capacitance and shunt resistance of his equivalent circuit, we have

$$L = 2S_1\omega_0, \quad \dots\dots (2.15)$$

$$C = 1/2S\omega_0, \quad \dots\dots (2.16)$$

$$P = 2S_1\delta. \quad \dots\dots (2.17)$$

§ 3. THE MOTIONAL IMPEDANCE

We can now proceed to a discussion of the motional impedance of the galvanometer, but we may note before so doing that it is proposed to regard the circuit in the static condition as non-reactive. There is no loss of generality in this, because the dynamic reactance of the instrument is always so much greater than the static value that the inclusion of the latter in the discussion would be a mere algebraic *tour de force*.

Eliminating θ from equations (2.1) and (2.2), we find

$$e = \{R + k^2D/(\alpha D^2 + \beta D + \gamma)\}i. \quad \dots\dots (3.1)$$

It is evident that the second term in the bracket represents the contribution made to the effective impedance of the circuit by the moving galvanometer coil. This term we call the *motional impedance* (more precisely it is the motional impedance operator) and we denote it by I . Then

$$I = k^2D/(\alpha D^2 + \beta D + \gamma) \quad \dots\dots (3.2)$$

$$= 2S\omega_0D/\{(D^2 + \omega_0^2) + \delta\omega_0D\}, \quad \dots\dots (3.3)$$

using equations (2.4), (2.5), (2.8) and (3.1).

From (2.1) and (3.2) we find

$$\theta = \frac{I}{k} \cdot \frac{1}{D} i = \frac{I}{k} \int i dt, \quad \dots\dots (3.4)$$

while from (3.1) and (3.4) we have

$$\theta = \frac{1}{k} \cdot \frac{I}{I+R} \cdot \frac{1}{D} e = \frac{1}{k} \cdot \frac{I}{I+R} \int e dt. \quad \dots\dots (3.5)$$

Equations (3.4) and (3.5) present the equations of the moving-coil galvanometer in a novel and simple form which is of considerable value in circuit analysis. Equation (3.5) is of particular interest, because it holds the key to the problem of the behaviour of a galvanometer as a fluxmeter and shows at once that for flux measurement the condition is that the motional impedance shall dominate the static resistance of the circuit.

The expression $I/(I+R)$ may be regarded as the "response operator" of the instrument and its circuit, and may conveniently be rewritten as a function of D , ω_0 and σ , where $\sigma = S/R$ may be called the "degree of damping" of the instrument. We have

$$\frac{I}{I+R} = \frac{2\omega_0\sigma D}{D^2 + \omega_0^2 + \omega_0\sigma D(2\sigma + \delta)}. \quad \dots\dots(3.6)$$

While we shall discuss later the application to particular forms of time-variation of e , the form taken by I for sinusoidal variations, which is useful in connection with vibration galvanometers, is interesting. If ω be the frequency of e , then we have at once from (3.2)

$$I = \frac{2j\omega_0\omega S}{(\omega_0^2 - \omega^2) + j\omega\omega_0\delta},$$

where j is the operator rotating through a right angle. This may be written

$$I = \frac{2S\delta}{z^2 + \delta^2} + \frac{2jSz}{z^2 + \delta^2}$$

where $z = (\omega_0/\omega) - (\omega/\omega_0)$, showing the existence of a motional resistance term, R_M , and a motional reactance term, X_M , given by

$$R_M = 2S\delta/(z^2 + \delta^2); \quad X_M = 2Sz/(z^2 + \delta^2).$$

It is easy to show that the locus of the end of the vector $R_M + jX_M$ is a circle of diameter $2S/\delta$, passing through the origin and with its centre on the R axis. Except very near to resonance, we can disregard δ and then X_M assumes the very simple form $X_M = 2S/\{(\omega_0/\omega) - (\omega/\omega_0)\}$, and this quantity can, in fact, be taken in many problems as representing the motional impedance of the instrument.

§ 4. ILLUSTRATIVE EXAMPLES

Two of the most interesting examples of the application of the galvanometer equations in the form derived above are to be found in the calculation of the inherent "noise level" of a galvanometer (that is, the background deflection due to spontaneous voltage fluctuations in the circuit) and the calculation of the response of an amplifier using a galvanometer.

(i) Noise level

From equation (3.5) we may write

$$D\theta = \frac{1}{k} \frac{I}{I+R} e.$$

Here e will be the spontaneous E.M.F. of the circuit, and in any frequency range df we may write

$$|D\theta|^2 = \frac{1}{k^2} I + R \quad \dots\dots(4.1)$$

where mean square quantities are indicated by the bars.

By Nyquist's theorem (Moullin 1938, p. 13) we may put

$$\bar{e}^2 = 4RkTdf, \quad \dots\dots(4.2)$$

where k is Boltzmann's constant and T is the absolute temperature. We

therefore have, for the mean kinetic energy of the suspended system arising from spontaneous fluctuations in the range $d\omega$,

$$\begin{aligned} \frac{1}{2}\alpha|D\theta|^2 &= \frac{1}{2} \frac{\alpha}{k^2} I + R \\ &= \frac{4\omega^2\omega_0^2\sigma^2}{k^2(\omega^2 - \omega_0^2)^2 + 4\omega^2\omega_0^2\sigma^2} \cdot 4RkT \frac{d\omega}{2\pi} \\ &= \frac{2\omega_0\sigma kT}{\pi} \frac{\omega^2}{(\omega^2 - \omega_0^2)^2 + 4\omega^2\omega_0^2\sigma^2} d\omega, \end{aligned} \quad \dots\dots(4.3)$$

from (2.10) and (2.14).

The total energy, E , is thus given by

$$E = \frac{2\omega_0\sigma kT}{\pi} \int_0^\infty \frac{\omega^2}{(\omega^2 - \omega_0^2)^2 + 4\omega^2\omega_0^2\sigma^2} d\omega = \frac{1}{2}kT. \quad \dots\dots(4.4)$$

This may also be set equal to the mean potential energy, $\frac{1}{2}\gamma|\theta|^2$, so that

$$|\theta|^2 = kT/\gamma. \quad \dots\dots(4.5)$$

Using (2.13) we may write, for the mean deflection $\bar{\theta}$,

$$\bar{\theta} = \mu(\pi kT/S\tau)^{\frac{1}{2}}, \quad \dots\dots(4.6)$$

where $\tau = 2\pi/\omega_0$ is the free period of the instrument. This corresponds to an R.M.S. current \bar{i} given by

$$\bar{i} = \bar{\theta}/\mu = (\pi kT/S\tau)^{\frac{1}{2}}. \quad \dots\dots(4.7)$$

Ising (1926) has given a similar result (but restricted to the case of critical damping) deduced from considerations of the effect of Brownian movement of gas molecules on the suspension. As Moullin (1938, p. 217) has pointed out, we must suppose that the total background noise arises from the combination of the two effects, and we would expect therefore a background corresponding to $\bar{i}\sqrt{2}$, that is, to

$$\bar{i}\sqrt{2} = (2\pi kT/S\tau)^{\frac{1}{2}}. \quad \dots\dots(4.8)$$

Taking $k = 1.4 \times 10^{-16}$ ergs per degree and $T = 290$, we have

$$\bar{i}\sqrt{2} \sim 5 \times 10^{-7} (S\tau)^{-\frac{1}{2}} \text{ absolute amperes.} \quad \dots\dots(4.9)$$

(ii) Galvanometer amplifier

If we consider an amplifying system such as that described by Gall (1942, 1945), the essential circuit of which is shown in figure 1, the response can be calculated very readily by the present method.

Referring to figure 1, the galvanometer is represented by $I+r$, where r is its D.C. resistance, Z and Z_M are elements of a feedback circuit, and i_3 is the amplified current arising from the application of an E.M.F. e through a resistance P to the galvanometer.

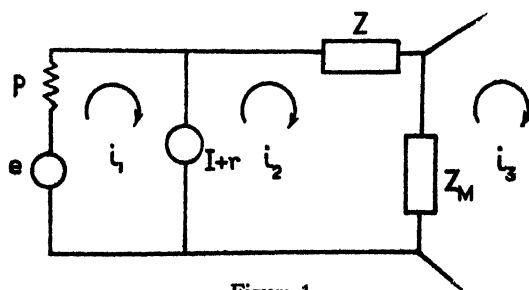


Figure 1.

We then have the following equations

$$\begin{aligned} e &= P i_1 + (I + r)(i_1 - i_2), \\ 0 &= (I + r)(i_2 - i_1) + Z i_2 + Z_M(i_2 - i_3), \\ i_3 &= g \theta ; \quad \theta = (I/kD)(i_1 - i_2), \end{aligned}$$

where g is the gain factor for the amplifying device. Solving for i_3 , we get for the current amplification

$$\frac{i_3}{i_1} = \frac{Z + Z_M}{Z_M} \cdot \frac{1}{1 + \xi} \quad \dots\dots(4.10)$$

where

$$\xi = kD(I + r + Z + Z_M)/gIZ_M. \quad \dots\dots(4.11)$$

The quantity ξ is a non-dimensional complex number. In any practical case it is easy to evaluate, and generally the two components are of the order of 0.01.

The effective input impedance, I' , of the galvanometer is given by

$$I' = (I + r) \frac{kD}{gI} \cdot \frac{Z + Z_M}{Z_M} \cdot \frac{1}{1 + \xi} \quad \dots\dots(4.12)$$

$$= (I + r) \frac{\text{Effective current gain}}{\text{Gain without feedback}}. \quad \dots\dots(4.13)$$

Thus in general $I' \ll I + r$.

In Gall's amplifier, Z and Z_M were pure resistances, Z/Z_M being of the order of 10^4 . If Z_M be of the form $S + MD$, i.e. a combination of mutual inductance M and resistance S , and if Z be a large resistance P , we have, disregarding ξ ,

$$i_3/i_1 = (P + S + MD)/(S + MD), \quad \dots\dots(4.14)$$

so that i_3 can be made to record the *integral* of i_1 over certain ranges of D . It is easy to make S/M about 0.001 (say with $M = 1\text{H}$, $S = 0.001\ \Omega$) and thus to secure an integrating time constant of the order of 1000 seconds (Astbury 1948).

(iii) Galvanometer response problems

The present calculus leads very readily to the operational method for the solution of certain response problems. We may take as simple examples the problem of determining the response to suddenly applied E.M.F. or flux signals. In the former case, let the E.M.F. be $H(t) \cdot e$, where $H(t)$ is the unit operator. Then from (3.5), we have

$$\theta = \frac{1}{k} \cdot \frac{I}{I + R} \cdot \frac{1}{D} H(t) \cdot e. \quad \dots\dots(4.15)$$

Writing $I/(I + R)$ as $f(D)$, we have (Jeffreys 1931)

$$\begin{aligned} \theta &= \frac{1}{k} \cdot f(D) \cdot \frac{1}{D} \cdot H(t) \cdot e \\ &= \frac{1}{k} \cdot \frac{e}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{\lambda t}}{\lambda^2} f(\lambda) \cdot d\lambda. \end{aligned} \quad \dots\dots(4.16)$$

For the second case, if the applied flux is $H(t) \cdot \Phi$, we have

$$\begin{aligned} \theta &= \frac{1}{k} \cdot \frac{I}{I + R} \cdot H(t) \cdot \Phi \\ &= \frac{1}{k} \cdot \frac{\Phi}{2\pi j} \int_{c-j\infty}^{c+j\infty} \frac{e^{\lambda t}}{\lambda} f(\lambda) d\lambda. \end{aligned} \quad \dots\dots(4.17)$$

The complex integral in both cases is taken round a curve from $c - j\infty$ to $c + j\infty$, c being positive and finite. The evaluation of the integrals is straightforward, since $f(\lambda)$ has simple poles at the roots of $\lambda^2 + 2\sigma\omega_0\lambda + \omega_0^2 = 0$.

(iv) *Two galvanometers in series*

Finally, we may note the easy way in which we can write down the solution to the problem of two galvanometers in series, an arrangement suggested by Butterworth for filtering. If I_1 and I_2 are the motional impedances of the instruments and R the static resistance of the circuit, the current i is given by

$$i = e / (R + I_1 + I_2) \quad \dots\dots (4.18)$$

and the deflections θ_1, θ_2 are given by

$$\theta_{1,2} = \frac{1}{k_{1,2}} \cdot \frac{I_{1,2}}{R + I_1 + I_2} \int e dt$$

and

$$\frac{I_{1,2}}{R + I_1 + I_2} \cdot \frac{2S_{1,2}/z_{1,2}}{\{R^2 + 4(S_1/z_2 + S_2/z_2)^2\}^{1/2}} \quad (4.19)$$

where the z s are as defined in (3.8).

ACKNOWLEDGMENTS

The writer is indebted to the Chief of the Royal Naval Scientific Service for permission to publish this paper, which formed part of work done during the war.

REFERENCES

- ASTBURY, N. F., 1948, in preparation.
 BUTTERWORTH, S., 1912, *Proc. Phys. Soc.*, **24**, 88.
 GALL, D. C., 1942, *J. Instn. Elect. Engrs.*, **89**, Pt. II, 434 ; 1945, *J. Sci. Instrum.*, **22**, 218.
 ISING, 1926, *Phil. Mag.*, Ser. 7, **1**, 827.
 JEFFREYS, H., 1931, *Operational Methods in Mathematical Physics* (Cambridge : University Press).
 MOULLIN, E. B., 1938, *Spontaneous Fluctuations of Voltage* (Oxford : Clarendon Press).

DISCUSSION

MR. J. F. W. BELL. I have recently been making use of the equivalent electrical circuit concept of the moving coil galvanometer as deduced by Butterworth. When a condenser is placed across the terminals there is an increase in the period, as would be expected if the galvanometer is looked upon as a parallel LCR network. The change in period gives the ratio k^2/γ . From a knowledge of the unloaded period and the current sensitivity the constants α , β and k can be calculated. This method of determining the constants of a galvanometer has given good agreement with nominal values.

There are possible applications of the galvanometer as an electromechanical resonator at power and low frequencies, similar to the use of the piezoelectric crystal resonator at radio frequencies. Its use in an electronically maintained sine wave generator is of particular interest.

MR. J. C. WILLMOTT. I would like to ask Mr. Astbury if the formula $\bar{\theta} = \mu \sqrt{\pi k T / s T_0}$ for the mean deviation is applicable whatever the resistance of the circuit connected across the galvanometer.

AUTHOR'S reply. I had originally intended to include as an example in the paper the problem quoted by Mr. Bell, and perhaps I may be allowed to show how simply the result can be obtained. If r be the D.C. resistance of the galvanometer, then the impedance, Z , obtained by connecting a capacitance C in shunt with the instrument, is given by

$1/Z = 1/(r + I) + j\omega C$ or $Z = (r + I)/[1 + j\omega C(r + I)]$. The maximum deflection will occur when $j\omega CI + 1 \approx 0$, from which $\omega^2 = \omega_0^2/(1 + 2SC\omega_0)$. Here ω_0 is the natural frequency of the instrument and ω the frequency as modified by the presence of the capacitance.

I feel that measurements of μ , S and ω_0 should be sufficient to provide estimates of all the other constants required, although I might mention that during the war I did devise a method for testing galvanometers based on a potentiometric method for measuring R_M and X_M at various frequencies. I agree with Mr. Bell that applications do exist for the galvanometer as he suggests, and I would like to add that it is good to see an appreciation of the galvanometer as a scientific instrument and not merely as a hack for null-detection work.

In reply to Mr. Willmott, the analysis given in the paper certainly leads to the result that the noise level is independent of R . We have made records of background signals which show that the level is substantially independent of R for values between critical resistance and one-fifth of this value, but the level observed was always higher by about 50% than that given by equation (4.9).

LETTERS TO THE EDITOR

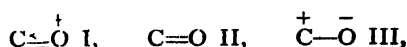
The Structure of Carbon Monoxide

In both the molecular orbital and pairing approximations it is easily shown that the four π electrons of carbon monoxide are completely equivalent, so that the distinction between bonding and non-bonding electrons recently introduced by Long and Walsh (1947) is invalid. The symmetry of the molecule is most simply exhibited by Mulliken's formulation (1932):

$$(K)(K)(\pi\sigma)^2(\gamma\sigma)^2(w\pi)^4(\sigma\sigma)^2, \quad {}^1\Sigma^+$$

where $(w\pi)$ is doubly degenerate.

The existence of the $(w\pi)^4$ shell, while indicating that four equivalent π electrons are concerned in $\pi\pi$ bonding, by no means implies that the molecule is best represented by a triple-bonded structure, as Long and Walsh seem to suppose. Indeed, analysis shows (Moffitt, to be published) that of the conventional structures



II is the most important, and I is less important than III.

The rotational analysis (Schmid 1932) of the $(\text{CO})^+$ produced by the first ionization of the molecule has shown it to have the symmetry ${}^2\Sigma$. Consequently the first ionization potential of CO, ${}^1\Sigma^+$ corresponds to the removal of a σ electron. The removal of a π electron, as postulated by Long and Walsh, would lead to $(\text{CO})^+$, ${}^2\Pi$, and Mulliken (1928) has cited evidence to show that this ionization occurs near 17 e.v. and not near 14 e.v.

It is difficult to follow Long and Walsh in their reasoning that the removal of a π electron should lead to a strengthening of the bonding in the resulting $(\text{CO})^+$. Introducing the wave-function symmetrization necessary to give significance to their structure, the ionization of an electron which accounts for a quarter of the considerable $\pi\pi$ bond strength in CO would surely lead to a weakening; and this is actually observed for CO^+ , ${}^2\Pi$ (Sponer 1935). The "three-electron bonding function" involving one π and two σ electrons, invoked in this connection, has, as far as I am aware, no precedent.

The low dipole moment of CO is well known; calculations (Moffitt, to be published) based on Long and Walsh's structure indicate, however, that the molecule should be highly polar. Further, the electron donor properties of CO are hardly to be attributed to the "inert pair" $(2s_e)^2$, whose ionization potential is expected to be near 20 e.v., as assumed by these authors.

The real difficulty in describing the carbon monoxide molecule does not lie in the specification of the LCAO forms for the $(w\pi)^4$ shell, but in characterizing the nature of the three different σ -type orbitals. A plausible assignment of the six σ electrons to three localized molecular orbitals has been devised (Moffitt, to be published). Retaining Mulliken's notation, $(\pi\sigma)^2$ now corresponds essentially to $(2s_e)^2$ and $(\gamma\sigma)^2$ refers to an electron pair

bond between some sp -type hybrid carbon orbital and a predominantly $2p\sigma$ oxygen orbital. The third pair of electrons $(x\sigma)^2$ are localized near the carbon atom and are nearly non-bonding. The ionization potential of a $(x\sigma)$ electron is estimated at near 14 eV., and, therefore, accounts for the first ionization potential and electron donor properties of carbon monoxide. The $(y\sigma)^2$ electron pair bond is shown to be largely homopolar.

The effect of the s - p hybridization gives rise to a compromise between strong $\sigma\sigma$ -type bonding and the promotion required to produce the appropriate carbon valence state. This compromise has been shown to lead to an asymmetry in σ electron distribution, whose direction is opposite and whose magnitude is approximately equal to that of the π electrons; the low dipole moment of carbon monoxide is no longer surprising.

The ionization of a non-bonding $(x\sigma)$ electron leads to a strengthening of the $\pi\pi$ bonding in as much as $(CO)^+$, $x^2\Sigma$ now becomes more nearly triple-bonded in the conventional sense. Further, the properties of the excited $A^1\Pi$ state of CO (Sponer 1935), in particular its weak bond strength, are accounted for by the formulation

$$(K)(K)(x'\sigma)^2(y'\sigma)^2(x'\pi)^4(x'\sigma)(\overline{y'\pi}), \quad {}^1\Pi,$$

where $\overline{(y'\pi)}$ is the antibonding orbital corresponding to $(y'\pi)$. A full account of this work is being prepared for publication.

The author would like to acknowledge the receipt of a grant from the Board of the British Rubber Producers' Research Association.

New College, Oxford.
5 March 1948.

WILLIAM MOFFITT.

- LONG J. H., and WAISH, A. D., 1947, *Trans. Faraday Soc.*, **43**, 342.
MULLIKEN, R. S., 1928, *Phys. Rev.*, **32**, 205; 1932, *Rev. Mod. Phys.*, **4**, 1.
SCHMID, R., 1932, *Phys. Rev.*, **42**, 182.
SPONER, H., 1935, *Molekulspektren* (Berlin: Springer).

A Note on the Shape of the Polepieces of a Synchrotron Magnet

The Birmingham synchrotron magnet has recently been described (Oliphant, Gooden and Hide 1947), and an interesting problem connected with its magnetic field has arisen. For certain reasons it is desirable that the magnetic field strength H along the central radius of the gap between the polepieces of a synchrotron ring magnet should, in any normal section of the ring, be related to the radius vector r drawn from the centre of the ring by an equation of the form $H \propto r^{\frac{1}{2}}$. What shape should the polepieces have to give such a field? If the direct field due to the current in the coils be neglected, an approximate solution of the problem can be obtained by regarding it as a two-dimensional one. This is justified by the large mean radius of the ring, which is 450 cm. in the Birmingham magnet.

Once the method of conformal transformations has been adopted, the fact that the integral of $r^{-\frac{1}{2}} dr$ is equal to $3r^{\frac{1}{2}}$ suggests the trial of

$$W = Az^{\frac{1}{2}} \quad \dots \dots (1)$$

as a possible solution. A is a constant and W is the complex potential of the form $W = U + iV$ and $z = x + iy = re^{i\theta}$. Applying equation (1), we get

$$W = Ar^{\frac{1}{2}}e^{i\theta/3} = Ar^{\frac{1}{2}}\cos \theta/3 + iAr^{\frac{1}{2}}\sin \theta/3.$$

Take V as the potential and U as the stream function, then, since

$$U = Ar^{\frac{1}{2}}\cos \theta/3; \quad V = Ar^{\frac{1}{2}}\sin \theta/3,$$

equipotentials are given by

$$r^{\frac{1}{2}}\sin \theta/3 = B \quad \text{or} \quad r = C \operatorname{cosec}^2 \theta/3, \quad \dots \dots (2)$$

and lines of force are given by

$$r^{\frac{1}{2}}\cos \theta/3 = D \quad \text{or} \quad r = E \sec^2 \theta/3. \quad \dots \dots (3)$$

Figure 1 represents sets of such curves.

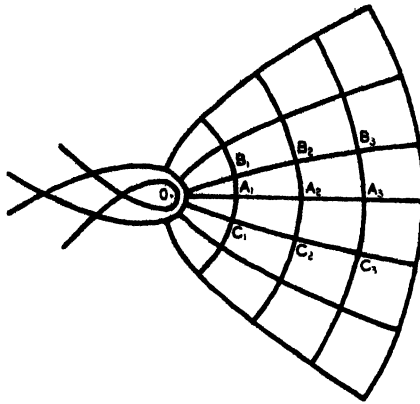


Figure 1. Section of field of "ideal" synchrotron magnet (right side only). Lines such as $B_1B_2B_3$, $C_1C_2C_3$ are equipotentials and possible polepiece profile curves. Lines such as $B_1A_1C_1$, $B_2A_2C_2$ are lines of force. Beams of particles pass normal to the plane of the figure through points on $A_1A_2A_3$.

The radial field strength $H_r = -\partial V/\partial r = -(Ar^{-1/3} \sin \theta/3)/3$, and the tangential component $H_\theta = -\partial V/r\partial \theta = -(Ar^{-1/3} \cos \theta/3)/3$. The resultant field strength is $H = Ar^{-1/3}/3$. Along the central radius $\theta=0$; thus along that line $H_r = 0$, $H_\theta = H = Ar^{-1/3}/3$. Hence the field satisfies the desired condition. At any point $H_r/H_\theta = \tan \theta/3$, showing that for small values of θ , H_r is very small compared with H_θ .

The polepieces are to be made with profile curves $r = C' \operatorname{cosec}^3 \theta/3$, for they are themselves magnetic equipotentials.

The complete curves given by (2) and (3) consist of closed loops with "wings" extending to infinity. The two sets of curves are of the same shape but the lines of force are obtained by rotating the other set through 270° about an axis through the origin normal to the plane of the figure. The necessity of making the real polepieces finite in length instead of infinite increases the discrepancy between theory and experiment, but in the middle of the gap the difference should not be great.

The "basic" shape of the profile curves of the Birmingham synchrotron magnet, obtained empirically by the aid of the Peierls tank, appears to be very close to that suggested by equation (2).

University College, Nottingham.
23 February 1948.

N. DAVY.

OLIPHANT, M. L., GOODEN, J. S., and HIDE, G. S., 1947, *Proc. Phys. Soc.*, **59**, 666.

REVIEWS OF BOOKS

The A Priori in Physical Theory, by ARTHUR PAP. Pp. x+102. (New York : King's Crown Press, 1946.) 13s. 6d. net.

This book deals with a problem which, mainly through the assertions of the late Sir Arthur Eddington and Professor E. A. Milne, has in recent years become of much interest to physicists, namely the question whether or not the laws of nature are *a priori*, i.e. derivable by reason apart from experience and so not liable to falsification by experience. The author does not, however, attempt to solve the ultimate problem, but points out that in practice physical laws are often treated as "functionally" *a priori*; thus, if an experiment or observation appears to violate a law we sometimes do not abandon the law but regard the experiment as not providing a fair example of it. Thus, when Uranus appeared to violate

Newton's law of gravitation, the law was not changed but Uranus was held to be disturbed. Laws sufficiently well grounded can acquire a high degree of immunity in this way. It was on such grounds that Poincaré maintained that one could always regard space as euclidean by attributing any apparently non-euclidean qualities to fields of force or non-rectilinear propagation of light. The book is clearly and interestingly written, though one feels that the author tends to continue his argument after the point has been adequately explained.

With the main thesis it is impossible to disagree, for it simply records what every physicist knows to be true. It inevitably raises another question, however, which the author does not consider—namely, what determines the choice between abandonment of the law and rejection of the observation as irrelevant? This is a pertinent question because, for example, while the anomalous behaviour of Uranus did not upset Newton's law of gravitation, the anomalous behaviour of Mercury did so. The difference was not merely that Neptune was observed and "Vulcan" was not, for the law was retained in spite of the irregularities of certain stars to which were assigned purely hypothetical companions. It would be interesting to see how the author would deal with this extension of his problem.

Mr. Pap is evidently better acquainted with physics than some philosophers who write on physical theory, and his views command respect. There are a few slips, however, which might be worth mentioning. The principle of conservation of mass is not rendered invalid by the discovery that mass depends on velocity (p. 61); it is, in fact, in order to maintain that principle that we must make mass depend on velocity. Again, neither in Newton's nor in Einstein's language is it true to say that force means $m dv/dt$ (p. 65); it means $d(mv)/dt$. An injustice is done to Galileo on p. 85 where it is said that he regarded the string of his pendulum as having no mass. He considered the effect of its mass very interestingly in the *Dialogues Concerning the Two Great Systems of the World*. Finally (p. 22), the author, apparently following Lenzen, seems to think that there is a "vicious circle" in making the length of a standard measuring rod depend on its temperature and defining a temperature scale in terms of length—a circle which needs a series of "successive approximations" to break it. This is not so, however, for a *thermoscope* is sufficient to make the prescription of the standard measuring rod definite, leaving us free to choose whatever type of *thermometer* we wish. Nor, in fact, do we ever measure temperature in terms of *measure*s of length; the marks on our glass tube containing mercury are arbitrary so far as the unit of length measurement is concerned, which is why it is possible (though not necessary) to regard length and temperature as independent fundamental measurements in dimensional theory.

HERBERT DINGLE.

Five-figure Tables of Natural Trigonometrical Functions, prepared by H.M. Nautical Almanac Office. Pp. 124+iv. (London: H.M. Stationery Office, 1947.) 15s. net.

Tables of the natural values of the four trigonometrical functions: sine, tangent, cotangent, cosine, for every 10 seconds of arc, and an auxiliary table for the cotangent for every second of arc to $7^{\circ} 30'$.

A. C. S.

Logarithmic and Trigonometric Tables, by J. B. DALE. Pp. 42+vi. (London: Edward Arnold and Co., 1947.) 2nd Edition. 2s. 6d.

A new edition, completely reset in very clear form.

A. C. S.

Note. Details of binding cases will be given in the July issue of the *Proceedings*.

INDEX

	PAGE
Alkali halides, polarization of second order Raman effect in	498
Ammonia, collision broadening of inversion spectrum	540
Ammonia, pressure broadening of inversion spectrum: disturbance of thermal equilibrium	83
Andrade, E. N. da C.: Device for maintaining constant stress in rod undergoing plastic extension	304
Arc discharges, transient, cathode spot behaviour in	424
Astbury, N. F.: Calibration of hydrophones and crystal transceivers	193
Astbury, N. F.: Moving-coil galvanometer as circuit element	590
Band overlap, and thermionic emission constants	360
Barkhausen effect	370
Bates, L. F., and Davis, J. H.: Effect of temperature on heat changes in magnetization of Ni-Si alloy	307
Bates, L. F., and Harrison, E. G.: Adiabatic temperature changes accompanying magnetization of ferromagnetic alloys	213, 225
Belin, R. E.: Radiosonde method for atmospheric potential gradient measurements	381
β -ray measurements, correction for self-weakening in	460
β -ray spectrum of ThC''D	466
Birks, J. B.: Measurement of permeability of low-conductivity ferromagnetic materials at cm. wavelengths	282
Blackman, M., and Michiels, J. L.: Efficiency of counting systems	549
Bleaney, B., and Penrose, R. P.: Collision broadening of inversion spectrum of ammonia: III. cross-sections for self-broadening and for mixtures with non-polar gases	540
Bleaney, B., and Penrose, R. P.: Paramagnetic resonance at low temperatures in chromic alum	395
Bleaney, B., and Penrose, R. P.: Pressure broadening of inversion spectrum of ammonia: II. disturbance of thermal equilibrium at low pressures	83
Bond lengths and electronic structure of coronene and pyrene	309
Book reviews	114, 212, 308, 398, 499, 599
Brewer, A. W., Cwilong, B., and Dobson, G. M. B.: Measurement of absolute humidity in extremely dry air	52
Broda, E., Grummitt, W. E., Guéron, J., Kowarski, L., and Wilkinson, G.: Correction for self-weakening in β -ray measurements	460
Broda, E., <i>see</i> Zajac, B.	
Bush, H. D., and Tebble, R. S.: Barkhausen effect	370
Calibration of hydrophones and crystal transceivers	193
Campbell, N. R., and Hartshorn, L.: Experimental basis of electromagnetism: II. electrostatics	27
Carbon monoxide, structure	597
Cathode spot behaviour in transient arc discharges	424
Cavity resonator method of measuring dielectric constant of polar liquids	71
Chaudhri, M., and Fenton, A. G.: Experiments with adjustable Geiger-Müller counters	183
Clegg, J. A., <i>see</i> Lovell, A. C. B.	
Collie, C. H., Hasted, J. B., and Ritson, D. M.: Cavity resonator method of measuring dielectric constants of polar liquids in centimetre band	71
Collie, C. H., Hasted, J. B., and Ritson, D. M.: Dielectric properties of water and heavy water	145
Collision broadening of inversion spectrum of ammonia	540
Contacts, electrical behaviour, influence of surface films	315

Corcoran, N., and Hough, J. M. : Computation of polar diagram of radio aerial, flat earth and vertical screen	203
Corrigenda	308
Cosmic-ray intensity, semi-diurnal variation	509
Coulson, C. A. : Excited electronic levels in conjugated molecules, u.v. absorption of naphthalene etc.	257
Coulson, C. A., <i>see</i> Duncanson, W. E.	
Coulson, C. A., <i>see</i> Moffitt, W. E.	
Counters, Geiger-Müller, adjustable, experiments with	183
Counting systems, efficiency of	549
Cwilog, B., <i>see</i> Brewer, A. W.	
Davis, J. H., <i>see</i> Bates, L. F.	
Davy, N. : Shape of polepieces of synchrotron magnet	598
δ -radiation, Kapitza's theory, extension	453
Désirant, M., and Shoenberg, D. : Penetration of magnetic field into superconductors, measurements on thin cylinders	413
Diagrams, solid, to illustrate resonance phenomena	132
Dielectric constant measurement, of polar liquids by cavity resonator method	71
Dielectric films, electric strength	243
Dielectric properties of water and heavy water	145
Dilworth, C. C. : Influence of surface films on electrical behaviour of contacts	315
Disintegration of nitrogen by deuterons, neutrons emitted in	523
Dissociation energy of NO molecule	533
Dobson, G. M. B., <i>see</i> Brewer, A. W.	
Donovan, B. : Specific heat of linear ionic lattice	325
Duncanson, W. E., and Coulson, C. A. : Electron momenta in atoms	175
Efficiency of counting systems	549
Eigenvalue problems, solution of	481
Electric strength of dielectric films	243
Electromagnetism, experimental basis : electrostatics	27
Electron bunching for msec. pulse generation	397
Electronic levels, excited, in conjugated molecules : symmetry and multiplicity of molecular states	270
Electronic levels, excited, in conjugated molecules : u.v. absorption of naphthalene	257
Electronic structure and bond lengths of coronene and pyrene	309
Electron momenta in atoms	175
Electron trap mechanism of luminescence in sulphide and silicate phosphors	574
Electrostatics—experimental basis of electromagnetism	27
Emission, thermionic, constants, and band overlap	360
Energy levels in insulators, and work function	13
Energy levels, in oxide cathode coatings	22
Evans, W. M., <i>see</i> Pack, D. C.	
Feather, N. <i>see</i> Zajac, B.	
Fenton, A. G., <i>see</i> Chaudhri, M.	
Ferretti, B., and Krook, M. : Solution of scattering and related problems	481
Fisher, D. G. : Molecular structure and arrangement in stretched natural rubber	99
Frank, F. C. : Isotopic abundance rule, and origin of nuclei	211
Fromme, K. D. : Rate of growth of current and behaviour of cathode spot in transient arc discharges	424
Galvanometer, moving-coil, as circuit element	590
γ -radiation from polonium, study of	501
Garlick, G. F. J., and Gibson, A. F. : Electron trap mechanism of luminescence in sulphide and silicate phosphors	574
Gerö, L., and Schmid, R. : Dissociation energy of NO molecule	533

Gibson, A. F., <i>see</i> Garlick, G. F. J.	
Gibson, W. M., and Livesey, D. L. : Neutrons emitted in disintegration of nitrogen by deuterons	523
Gold in silver, heat of solution and residual resistance	161
Grummitt, W. E., <i>see</i> Broda, E.	
Guéron, J., <i>see</i> Broda, E.	
Harrison, E. G., <i>see</i> Bates, L. F.	
Hartshorn, L., <i>see</i> Campbell, N. R.	
Hasted, J. B. : Milli-microsecond pulse generation by electron bunching	397
Hasted, J. B., <i>see</i> Collie, C. H.	
Heavy water, and water, dielectric properties	145
Helium, liquefaction, expansion method	405
Horton, G. K. : Angular distribution in internal pair creation	457
Hough, J. M., <i>see</i> Corcoran, N.	
Hsü, Yun-Kuei, <i>see</i> Martin, D. G. E.	
Huang, Kun : Quantum mechanical calculation of heat of solution and residual resistance of gold in silver	161
Humidity, absolute, measurement in extremely dry air	52
Hydrophones, and crystal transceivers, calibration	193
Impedance measurement, Lecher wire method	388
Insulators, energy levels in, and work function	13
Internal pair creation, angular distribution in	457
Isotopic abundance rule, and origin of nuclei	211
James, H. J., <i>see</i> Pack, D. C.	
Kapitza's theory of δ -radiation, extension	453
Kikuchi-line patterns, interpretation and application : I. crystal unit cell	341
Kowarski, L., <i>see</i> Broda, E.	
Krook, M., <i>see</i> Ferretti, B.	
Kun Huang, <i>see</i> Huang	
Lattices, ionic, linear, specific heat of	325
Lead sulphide and lead selenide photosensitive deposits. structure, and effect of sensitization by oxygen	117
Livesey, D. L., <i>see</i> Gibson, W. M.	
Loeb, L. B. : Certain aspects of mechanism of spark discharge	561
Longuet-Higgins, H. C. : Excited electronic levels in conjugated molecules : symmetry and multiplicity of molecular states	270
Lovell, A. C. B., and Clegg, J. A. : Characteristics of radio echoes from meteor trails : I. intensity and electron density	491
Luminescence, electron trap mechanism, in sulphide and silicate phosphors	574
Magnetization of ferromagnetic alloys, adiabatic temperature changes accompanying	213, 225
Magnetic field, penetration into superconductors	413
Magnetization of Ni-Si alloy, effect of temperature on heat changes	307
Magnetization, variation of Young's modulus for Co with	236
Martin, D. G. E., Richardson, H. O. W., and Hsü, Yun-Kuei : β -ray spectrum of $\text{ThC}''\text{D}$	466
Menzies, A. C., and Skinner, J. : Polarization of second order Raman effect in alkali halides	498
Metals, slip at grain boundaries and grain growth	391
Meteor trails, characteristics of radio echoes from : I. intensity and electron density	491
Michiels, J. L., <i>see</i> Blackman, M.	
Moffitt, W. : Structure of carbon monoxide	597

	PAGE
Moffitt, W. E., and Coulson, C. A. : Electronic structure and bond lengths of coronene and pyrene	309
Molecular states, symmetry and multiplicity	270
Molecular states, and u.v. absorption of naphthalene etc.	257
Molecular structure and arrangement in stretched natural rubber	99
Mott, N. F. : Slip at grain boundaries, and grain growth in metals	391
Moving-coil galvanometer as circuit element	590
Nicolson, P., and Sarabhai, V. : Semi-diurnal variation in cosmic-ray intensity	509
Nitrogen, disintegration by deuterons, neutrons emitted in	523
NO molecule, dissociation energy	533
Nuclear physics, internal pair creation, angular distribution	457
Optical instruments, perfect	293
Oxide cathode coatings, energy levels in	22
Pack, D. C., Evans, W. M., and James, H. J. : Propagation of shock waves in steel and lead	1
Paramagnetic resonance, in chromic alum, at low temperatures	395
Penrose, R. P., <i>see</i> Bleaney, B.	
Permeability of low-conductivity ferromagnetic materials at cm. wavelengths, measurement	282
Phosphors, sulphide and silicate, electron trap mechanism of luminescence	574
Photosensitive PbS and PbSe deposits, structure	117
Pickard, G. L., and Simon, F. E. : Quantitative study of expansion method for liquefying helium	405
Plessner, K. W. : Electric strength of dielectric films	243
Polar diagram of radio aerial, flat earth and vertical screen, computation of	203
Polar liquids, dielectric constant measurement by cavity resonator method	71
Polonium, study of γ -radiation from	501
Potential gradient, atmospheric radiosonde method of measurement	381
Pressure broadening of inversion spectrum of ammonia : II. disturbance of thermal equilibrium at low pressures	83
Prowse, W. A. : Solid diagrams illustrating resonance phenomena	132
Pulse generation, milli-microsecond, by electron bunching	397
Quantum mechanical calculation of heat of solution and residual resistance of gold in silver	161
Radiosonde method for atmospheric potential gradient measurement	381
Raman effect in alkali halides, polarization	498
Reflectivity, metallic, variation with temperature	8
Resonance, paramagnetic, at low temperatures in chromic alum	395
Resonance phenomena, solid diagrams illustrating	132
Richardson, H. O. W., <i>see</i> Martin, D. G. E.	
Ritson, D. M., <i>see</i> Collie, C. H.	
Rubber, natural, stretched, molecular structure and arrangement in	99
Rubber, stresses and birefringence in, when subjected to strain	135
Sarabhai, V., <i>see</i> Nicolson, P.	
Scattering problems, solution of	481
Schmid, R., <i>see</i> Gerö, L.	
Self-weakening in β -ray measurements, correction for	460
Shoenberg, D., <i>see</i> Désirant, M.	
Shock waves, propagation in steel and lead	1
Sillitto, R. M. : Extension of Kapitza's theory of δ -radiation	453
Simon, F. E., <i>see</i> Pickard, G. L.	

	PAGE
Sinha, S. P.: Blue and ultra-violet bands of K_2	436
Sinha, S. P.: Ultra-violet bands of Li_2	443
Sinha, S. P.: Ultra-violet bands of NaK	447
Skinner, J., <i>see</i> Menzies, A. C.	
Slip at grain boundaries and grain growth in metals	391
Smith, T.: Perfect optical instruments	293
Spark discharge, mechanism of	561
Specific heat of linear ionic lattice	325
Spectrum, inversion, of ammonia, collision broadening	540
Spectrum, inversion, of ammonia, pressure broadening	83
Spectra, blue and ultra-violet bands of K_2	436
Spectra, ultra-violet bands of Li_2	443
Spectra, ultra-violet bands of NaK	447
Street, R.: Variation with magnetization of Young's modulus for cobalt	236
Stress, constant, maintenance in rod undergoing plastic extension	304
Stresses and birefringence in rubber subjected to strain	135
Superconductors, penetration of magnetic field: measurements on thin cylinders	413
Surface films, influence on electrical behaviour of contacts	315
Synchrotron magnet, shape of polepieces	598
Tebble, R. S., <i>see</i> Bush, H. D.	
Temperature changes, adiabatic, accompanying magnetization of ferromagnetic alloys	213, 225
Temperature, variation of metallic reflectivity with	8
ThC'D, β -ray spectrum of	466
Transceivers, crystal, and hydrophones, calibration	193
Treloar, L. R. G.: Stresses and birefringence in rubber subjected to homogeneous strain	135
Ultra-violet bands of K_2 , Li_2 , NaK	436, 443, 447
Water, and heavy water, dielectric properties	145
Weil, R.: Variation with temperature of metallic reflectivity	8
Wilkinson, G., <i>see</i> Broda, E.	
Williamson, M.: Lecher wire method of measuring impedance	388
Wilman, H.: Interpretation and application of electron-diffraction Kikuchi-line patterns: I. crystal unit cell	341
Wilman, H.: Structure of photosensitive PbS and PbSe deposits and effect of sensitization by oxygen	117
Wohlfarth, E. P.: Thermionic emission constants and band overlap	360
Work function, and energy levels in insulators	13
Wright, D. A.: Energy levels in oxide cathode coatings	22
Wright, D. A.: Work function and energy levels in insulators	13
Young's modulus for cobalt, variation with magnetization	236
Zajac, B., Broda, E., and Feather, N.: Further study of γ -radiation from polonium	501

INDEX TO REVIEWS OF BOOKS

	PAGE
Boucher, Paul E. : <i>Fundamentals of Photography with Laboratory Experiments</i>	402
Bouwers, A. : <i>Achievements in Optics</i>	212
Chase, C. T. : <i>The Evolution of Modern Physics</i>	403
Cork, J. M. : <i>Radioactivity and Nuclear Physics</i>	400
Cosslett, V. E. : <i>The Electron Microscope</i>	404
Culver, Charles A. : <i>Theory and Applications of Electricity and Magnetism</i>	499
Dale, J. B. : <i>Logarithmic and Trigonometric Tables</i>	600
Evans, David T. : <i>Frontiers of Astronomy</i>	404
Griffiths, Ezra : <i>Methods of Measuring Temperature</i>	308
Lovell, Bernard (Editor) : <i>Electronics and their Applications in Industry and Research</i>	114
Mathematical Tables Project of the National Bureau of Standards : <i>Table of the Bessel Functions $J_0(x)$ and $J_1(x)$ for Complex Arguments</i>	401
Mathematical Tables Project of the National Bureau of Standards : <i>Tables of Spherical Bessel Functions</i>	401
Nautical Almanac Office : <i>Five-figure Tables of Natural Trigonometrical Functions</i>	600
Nimmo, R. R. : <i>Atomic Energy</i>	403
Pap, Arthur : <i>The A Priori in Physical Theory</i>	599
Poynting, Thomson and Todd : <i>A University Text-Book of Physics, Volume I: Properties of Matter</i>	404
Ridenour, Louis N. (Editor) : <i>Radar System Engineering</i>	398
Snoek, J. L. : <i>New Developments in Ferromagnetic Materials</i>	116
Ubbelohde, A. R. : <i>Time and Thermodynamics</i>	402

